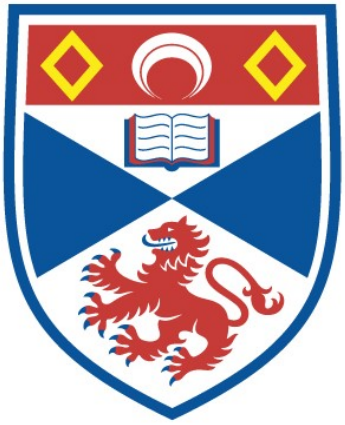
TESTING THE PREDICTIVE CODING ACCOUNT OF TEMPORAL INTEGRATION IN THE HUMAN VISUAL SYSTEM- A COMPUTATIONAL AND BEHAVIOURAL STUDY

Fraser Aitken

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2019

Testing the Predictive Coding Account of Temporal Integration in the Human Visual System-A Computational and Behavioural Study

Fraser Aitken

University of
St Andrews

A Thesis Submitted for the Degree of PhD at the University of St Andrews

October 2019

Funding

This work was supported by the University of St Andrews Wellcome trust strategic support grant [grant code APSC-ISSFFA].

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at https://osf.io/dashboard & http://research-repository.st-andrews.ac.uk/

Candidate's declaration

I, Fraser Aitken, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 68,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also

understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Fraser Aitken, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date Signature of candidate

Date Signature of supervisor

Abstract

A major goal of vision science is to understand how the visual system maintains behaviourally relevant perceptions given the level of uncertainty in the signals it receives. One proposed solution is that the visual system applies predictive coding to its inputs based on the integration of prior knowledge and current stimulus features. However, support for some vital aspects of predictive coding in the temporal domain is lacking and simpler accounts of temporal integration also exist. The aim of this thesis was to test two key attributes of predictive coding in time a) does the visual system apply adaptive weighting to prediction errors and b) can the visual system apply probalistic information learnt from stimulus sequences when making predictions. In chapters 3 & 4, we tested predictive coding's ideas of how prediction errors are weighted under the theoretical guidance of a temporal integration model linked to predictive processing, called the Kalman filter. Here, both experiments supported predictive coding. We showed that, consistent with the Kalman filter, visual estimates and the way estimation errors were corrected, adapted to stimulus behaviour and viewing conditions. In chapter 5, we assessed the ability of the visual system to integrate conditional relationships present in sequences of stimuli when making predictions. To do this, we inserted a stimulus sequence that changed and omitted trials based on Markov transition probabilities that made some transitions more or less probable and assessed reaction times and omission trial responses. Reaction time data was consistent with predictive coding, in that more predictable changes elicited faster responses. Omission trials data, was though, less clear. When faced with no stimulus, participants did not apply the conditional probabilities in their decisions optimally, instead applying non optimal decision strategies, inconsistent with predictive coding. In summary, this thesis supports the predictive coding of temporal integration but questions its application in all situations.

General acknowledgements

I would first like to thank my supervisor Dr Justin Ales for his guidance and support throughout my PhD. I will be forever grateful to him for providing the opportunity to study at such a fabulous University and extend this gratitude to the Wellcome trust strategic fund for providing financial support. An unforgettable experience. The end of this PhD marks the end of 8 years as a student at one level or another. With this in mind, I would like to thank Alan Searle at the University of Hull for making me believe that I could progress in academia and Professor Alex Wade at the University of York for inspiring my interest in the visual system. I would also like to thank everyone in the Vision group at St Andrews, in particular Andy Mackenzie for his words of wisdom about negotiating a PhD. Pastorally, I would also like to extend thanks to Professor Julie Harris and Dr Ines Jentzsch for their chats and guidance during my time at St Andrews. Always happy to listen to students and provide useful advice when required. Of course, no acknowledgements list would be complete without mentioning the good folk of the Lilac room. Big thanks go to my friends Abigail and Alex, who have been 'in it' from the beginning with me since I arrived at St Andrews. No mention of the Lilac room would be complete without mentioning Alonso, Sonny, Gierdre, Jenny and especially Bashar and 'Charley' Charles Ogunbode for their dependable good humour and friendship. No doubt the Criterion will miss us all! I would also like to say thank you to Dr Peter Kok at UCL for allowing me the time to work on my thesis in my new job. I also give a big thanks to my Mum and Dad for their support since entering academia. Finally, the biggest thanks of all goes to my wonderful wife Karen. There were times when I doubted whether this day would ever come but Karen was always there to keep my spirits up and ensure I remained on track. I am truly lucky to have her in my corner and grateful to her for being there.

Table of contents

List of figures chapter 1

List of tables

List of important equations

List of figures

List of figures

List of tables.

List of figures

List of figures

List of tables

Chapter one. Vision and uncertainty: how to deal with a dynamic and noisy world.

1. General Introduction

1.1 Vision: a vital behavioural tool.

Nervous systems have evolved under natural conditions to extract and compute behaviourally important information from the external world (Zeil, Boeddeker, & Hemmi, 2008). Of all parts of the nervous system that computes such information the visual system is perhaps the most important. The reason why vision is so important, is that we use vision in practically every behaviour we perform. Crucial tasks, such as navigating our environment, detecting danger, finding food, using tools and successfully interacting with others, all rely heavily on vision. However, despite our heavy reliance on vision, as we go about our busy daily lives, in most cases we seldom give thought to how important vision is or how it might actually work. One reason for this indifference is that in most situations vision seems remarkably easy. In fact, vision, unlike other cognitive processes such as solving a word puzzle or a mathematical problem seems remarkably straightforward. No effort at all is really needed to produce a solution. All we need do is open our eyes and the world is there before us instantly as a constant, accurate and stable perception of the outside world of sufficient resolution and speed to facilitate effective behaviours. Nonetheless, this apparent expertise in perceiving the world and the ease in which we can guide behaviours using vision belies a task of true complexity for the visual cortex.

1.2. The central problem for vision.

The crux of the problem facing the visual system is that the visual cortex, in which perception actually 'exists', is firmly encased within the skull. This means that the visual cortex does not have direct access to stimuli of interest or events in the external world. Isolation from the world presents an inverse problem for the visual system in that it must reliably interpret and adapt to unknown events and objects in the environment based only on indirect information (Pizlo, 2001). The source of this indirect information is light that reflects or emits from stimuli in the world forming information bearing patterns on the retinal surface (Hubel, 1988). Relying on external light signals and the patterns they form on the retina to internally interpret the world opens up a myriad of problems for the visual cortex which combine to make visual information uncertain (Gregory, 1970; Gordon et al, 2017; Knill & Pouget, 2011.; Kwon, Tadin, & Knill, 2015; Levi, Klein, & Chen, 2005; Parr, Rees, & Friston, 2018; Wei, Wert, & Körding, 2010). The notion of uncertainty is prominent in vision science, thanks in no small part to the influence of Swets & Green (1967) and their signal detection theory, plus a large body of prominent work into the way sensory systems reconcile noisy and ambiguous visual information (Beierholm, Quartz, & Shams, 2009; Casini, Mckay Illari, Russo, & Williamson, 2011; Gia Thu & Loc Hung, 2003; Knill, 2007; Knill & Pouget, 2004; Seth, 2014; Yuille & Kersten, 2006)

to name but a few. However, despite the widespread nature of research into how the visual cortex resolves uncertainty and the use of the term, exactly why visual information is uncertain is often not fully explained.

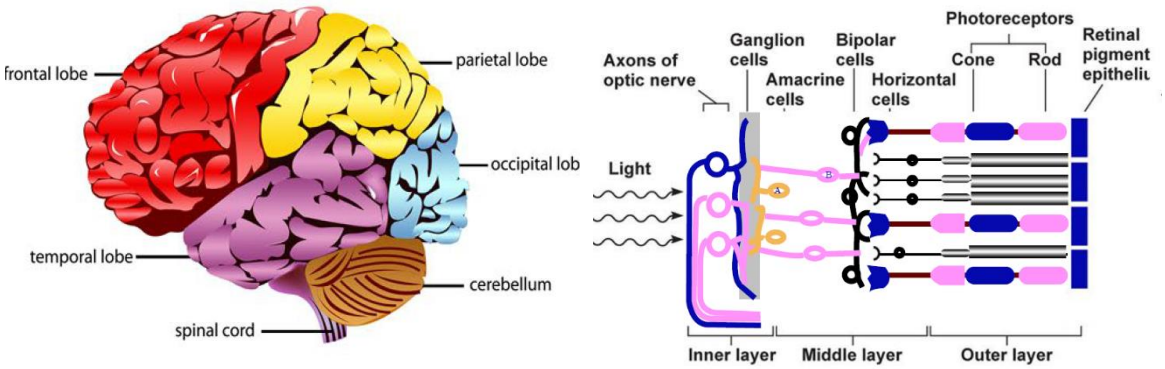


Figure 1. An illustration of the position of the isolated occipital cortex at the posterior of the brain source and the internal structure of the retina which provides visual information to the cortex (Figures A & B adapted from Hubel (1988)).

1.3. Uncertainty in visual information: variability in visual inputs.

One way of understanding why visual information is uncertain is to think about the flow of information the cortex receives from the retina as carrying statistics adapted to the external world in some way (Barlow, 1961; Berry, Warland, & Meister, 1997; Ly & Doiron, 2017) and perception as an interpretation of this information. The problem for the visual cortex is that like any system applying statistical interpretations to indirect signals from the outside world, the interpretation of signals arising from stimuli in the world is never entirely certain (Gregory, 1970; Yuille & Kersten, 2006). This is because all sensory information provided to the cortex from retinal measurements of the external world and early visual systems is to some extent variable (Knill & Pouget, 2004; Wolpert, Ghahramani, & Jordan, 1995; Wolpert, 2007). In the same way that increased variability in experimental measurements makes interpretation less certain (Taylor, 1997) (see figure 2 below), it also makes the interpretation of information from stimuli in the world more uncertain (Knill, 2007; Kwon et al., 2015). Due to the complexity and behaviour of the world and the way the retina and early visual systems behave there are a number of sources of variability present in visual information received at the cortex.

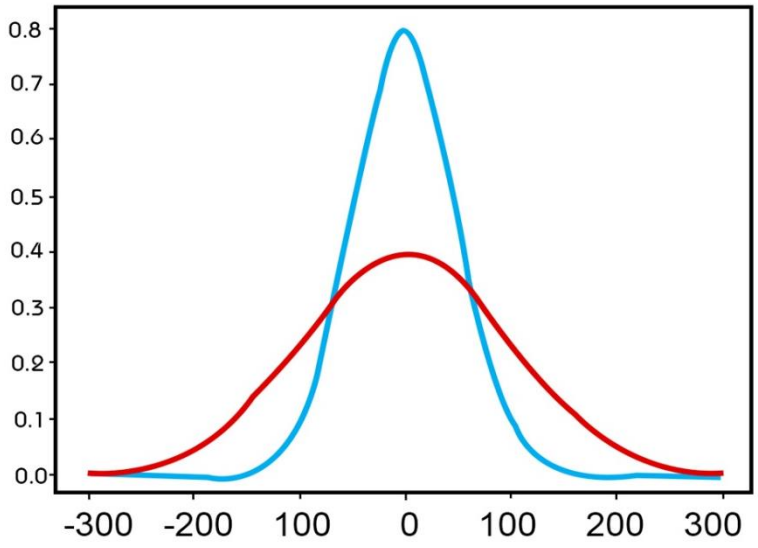


Figure 2. The general relationship between variability and uncertainty. Uncertainty simply means that we do not know the value of some quantity or outcome of an event, for example the current state of a changing stimulus based on an indirect noisy retinal image. The level of variability refers to the relative stability of values that we have recorded about the stimulus we are interested in. If we make lots of recordings that are around the same value then the variance in these recordings is low. In this case, if we plot the distribution of the data points we will have a narrow distribution (see the blue distribution) and we can attach a good level of certainty that the current recording of our stimulus if it falls within this distribution. Alternatively, if we make recordings that change considerably over time our recordings will carry a higher level of variability and our distribution will be broader (red distribution). This means that we can have less confidence that our current recording of our stimulus value is accurate.

1.4. Sources of variability in visual information.

The combination of the behaviour of the external natural world and the workings of human anatomy means that there are a number of sources of variability present in the visual information that is interpreted by the brain (Burge, Ernst, & Banks, 2008; Melcher, 2011; Wei et al., 2010). Importantly, the nature of these sources of variability mean that they have different implications for perception which can be quite subtle (Burge et al., 2008; Burge, Girshick, & Banks, 2010). To help simplify the sources of variability and their significance for perception, we split them into two groups. These are external sources of variability from events and stimuli outside of the brain and sources of internal variability that comes from the workings of the brain itself.

1.5. External variability

The first type of external variability present in incoming signals is simple. This is variability caused by the behaviour of objects and events in the world. When our surroundings are stable, visual information pertaining to stimuli is less variable and when our surroundings are changing it is more variable. This means that the amount of variability present in visual information can potentially provide important cues about what is happening in the external world. In an ideal world, the signals produced by stimuli under both changing and stable conditions would be easy to interpret. Small levels of variability should be taken to mean stability and high levels of variability would mean change has occurred and we should update our interpretation and perceptions accordingly (Denève, Duhamel, & Pouget, 2007; Wolpert & Flanagan, 2001). However, in the stream of visual information from the world there are other sources of external variance that do not arise from stimuli and their behaviours and add unwanted variability to the incoming stimulus information that we receive.

External variability can often arise from the environmental conditions in which we view the world. For example, the local atmospheric conditions through which light signals pass through on the way to the retina (See Saleh & Teich (2001) for a comprehensive account of the behaviour of light in different atmospheric conditions). If the weather is clear, then light signals and the information they contain travel to the retina largely intact (Jägerbrand & Sjöbergh, 2016). In this situation, extraction and interpretation of the part of the signal relating the behaviour of the stimuli of interest is more straightforward. However, the weather is not always clear. In situations such as fog or rain, the level of variability in light signals can be increased or the part of the signal carrying stimulus information degraded or lost. This is because light changes its behaviour as it travels between lighter and denser mediums such as air and water. When light carrying stimulus information from a car or road sign passes thorough a water droplet in a rain shower or low cloud formation (see figure 3 below) only a portion of the light from the sign passes through to the retina while the rest scatters. Weather conditions can therefore block or deflect some of the information from stimulus in the world and prevents the cortex from accessing as much stimulus information that would otherwise be available.

Figure 3. Effects of poor weather on viewing conditions. Here we see the how the effects of poor weather degrade the quality of signals carrying task relevant information. When driving we need to know about the behaviour of other cars on the road. However, here water in the atmosphere has blocked some of the light to while diffused other parts of the light carrying important signals (image is my own).

Visual information can also become more variable due to light changes during different phases of the diurnal cycle. While artificial light is relatively abundant in today's society the vast majority of the basis for signals from external stimuli still comes from direct or indirect light from the sun. As the suns elevation declines from it midday peak (60-90°) to the horizon (0°) at sunset, light intensity declines approximately 100 fold with the majority of this drop occurring rapidly in the final 5° of decline (Warrant & Johnsen, 2013) The effect of reduced light is to produce a gradient decline in the intensity of the light carrying stimulus information measured at the retina as the amount of light in the environment declines (Jägerbrand & Sjöbergh, 2016). The decline in light has an inverse relationship with variability in retinal measurements because as when light levels go down, variability in the measurement at the cortex goes up (Cordani et al., 2018) and because the sun sets and rises every day represents a twice daily source of additional variability in visual information from stimuli in the world (see figure 4 below).

Figure 4. Effects of reduced light on viewing conditions. Here we can observe a degraded image due to failing light. In this case we have a driving in twilight example. We see clear reduction in object contrast and spatial frequency due to the decreased light (image is my own).

External variability is also caused by the workings of human anatomy. One such anatomically related source of variance is indirectly caused by the structure and organisation of photoreceptors in the retina. The retina contains two types of two types of photoreceptors. Namely, Rods and Cones. Rods are more numerous, some 120 million, while there are only 6 million cones (see Solomon & Lennie, (2007) for a descriptive account of the structure and function of retinal photoreceptors). However, while cones are less numerous, they are more tightly packed together in the central fovea of the retina. In much the same way increased pixel density provides higher resolution in a television images, the high density of cones also promotes high resolution spatial and colour vision in the centre of our visual field (Hubel, 1988). The balance between rods and cones and their respective positions and densities has proven effective in general terms but it has not come without some cost. Specifically, that in order to focus our high resolution spatial and colour central vision on task relevant stimuli, we need to constantly move our eyes, head and body to some extent. This means that even during fixation our eyes are in motion. Almost constant anatomical motion means that retinal patterns are rarely stable on the photoreceptors of the eye (Arathorn et al, 2013; Melcher, 2011). The upshot of instability in information bearing patterns on the retina is to introduce motion related variability into the image of stimuli in the world and is an almost constant source of uncertainty in visual information.

A further source of anatomically related external variability is caused by the need to maintain a moving eye. A large part of this maintenance is performed by blinks. Blinks are an essential function of the eye that help spread tears across and remove irritants from the surface of the cornea and conjunctiva (Hall, 1945). Of course blinks are vital but there is some trade-off between the maintenance of the eye and the flow of visual information to the brain. A good way to understand this trade-off is to think of blinks as an on and off switch in the flow of stimulus information to the cortex. When our eyes are blink free the flow is 'on' and during a blink the flow is 'off'. Clearly, blinks are short in duration but they occur very frequently at a rate of approximately 15 times every waking minute (Burr, 2005). Due to this frequency, blinks are a very common source of variability in information received by the visual cortex as they produce almost constant gaps in the flow of visual information.

1.6. Internal variability.

Internally produced variability comes from the workings of the brain in the form of neural noise (Stein, 2005). Neural noise is perhaps the most intriguing of all of the sources of variability present in visual information received at the cortex. This is due to the controversy over whether such variability should be considered as noise in the negative sense of the term (Averbeck, Latham, & Pouget, 2006). Neural noise is caused by the random electrical firings and fluctuations of neurons in the brain that do not appear to be related with encoding an external stimulus directly (Swain & Longtin, 2006). Previously, it was thought that neural noise served no benefit and was detrimental to sensory processing (Strong et al, 1998). More recently, it has been proposed that neural noise is actually beneficial to the computations the brain applies to interpreting uncertain inputs. One idea forwarded by proponents of Bayesian brain theories relates to the idea that the brain represents temporal information from the world as Poisson probability distributions (Averbeck, Latham, & Pouget, 2006). Bayesian computing is done most efficiently when both past information and current information carries the same type of probability distributions. It is proposed, that by adding noise to sensory information, neurons are trying to make incoming visual information as 'Poisson' as possible to match the internal representation of past information and thus maximise information transfer (Zylberberg et al, 2017). While this idea may or not be correct, we nonetheless include neural noise as a source of variability in visual information as it does add variability to information received at the cortex regardless of its function.

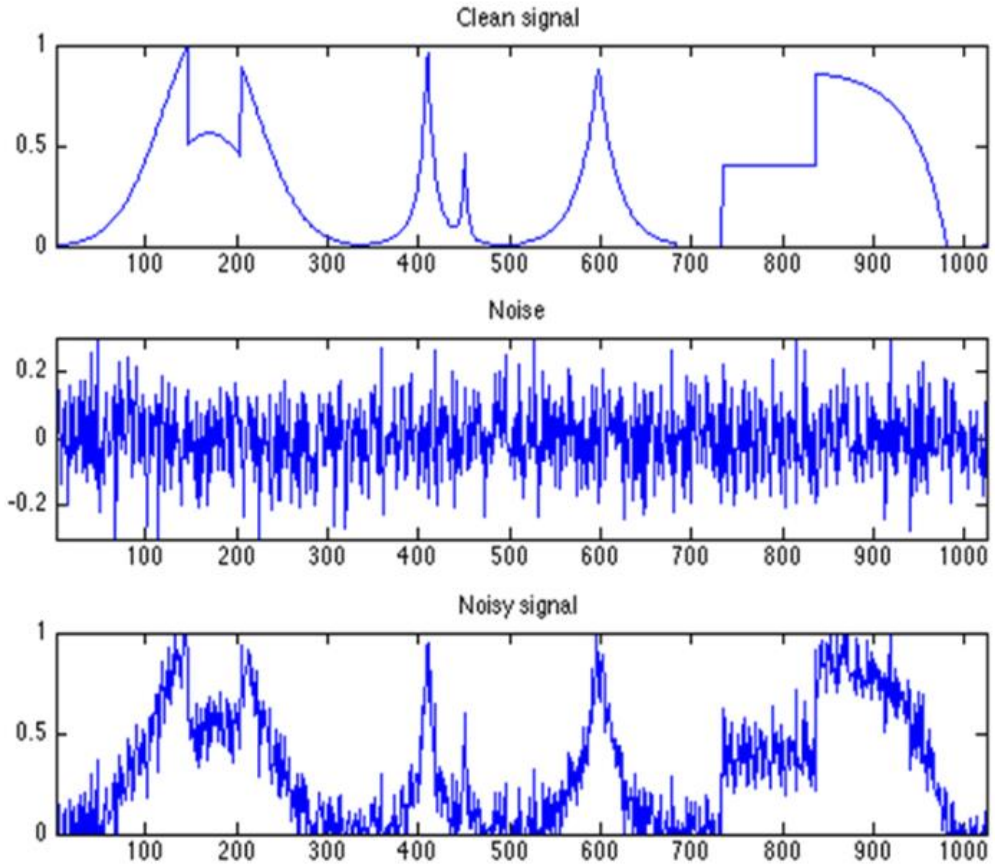


Figure 5. Variability, noise and stimulus information. Here using an example of radio signals we show the effects of factors such as the weather and neural noise might have on 'clean' stimulus information**.** Radio signals are an apt simile to visual signals as they are much the same as light signals. The only difference in the wavelength of the waves which are subject to many of the same sources of noise and variability. In the upper panel is a 'clean' radio signal. The variability in the frequency of this signal carries information about the content of the information transmitted by the sender. On its own interpretation of the clean signal is straightforward as the signal carries only variability about the signal of interest. In the middle panel we have only noise. Noise is unrelated to the radio signal carrying stimulus information. In radio waves, like light waves, noise can arise from atmospheric conditions or noise from other task irrelevant signals such as from other radio waves transmitting in a similar channel. It can also arise from components in the radio such as the flow of charges in the radio device (much like neural noise). The lower panel illustrates the sum total of signal and noise which must be interpreted by the receiving device. As we can see the addition of noise makes the radio signals of interest much more variable, which makes the true signal more uncertain and harder to interpret (figure adapted from signal and noise https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/9554/versions/5/previews/numerical-tour/denoising_noise_models/index.html accessed 1/10/16).

1.7. Variability in signals and measurements produces little effect on perception and behaviour: application of prior knowledge.

The internal and external sources of variability we have presented are certainly not trivial. It is apparent that the information that reaches the cortex about stimuli in the world is very often or most likely always variable making interpretation of important stimulus information constantly uncertain to some degree. Given the nature of the factors which produce uncertainty in visual information we might logically expect certain effects of different sources of variability to be present in perception. Because we constantly need to blink we might conclude that they would seriously interrupt our flow of visual consciousness. In addition, we might also expect our perception of objects in the world to be unstable nearly all of the time due to constant retinal instability or perception in bad weather or poor light to be worse than it actually is. Furthermore, we might think that neural noise would lead to confusion in the interpretation of incoming signals as it adds random fluctuations to the incoming information. However, in reality both internal and external variability appear to exert little effect on perception and behaviour in normal circumstances. We barely notice blinks despite occurring almost constantly, perception is remarkably stable at all times and our perception in bad weather and poor light although somewhat decreased in acuity is still normally reliable enough to perceive relevant stimuli and respond accordingly. Therefore, the contrast between the variable nature of visual information received at the cortex and our subjective visual and behavioural experience raises an important and currently unanswered question. Specifically, what computational processing strategies is the visual cortex applying to its inputs to extract relevant stimulus information from the mass of uncertain visual information emitting from world and turn this information into the high grade perceptions we are familiar with.

Exactly how the brain reconciles uncertainty is not fully understood. However, one strategy the visual system does appear to employ is to apply prior knowledge to interpreting current visual inputs (Friston, 2010; Kok, & De Lange, 2016). Importantly, the application of previous experience to current visual inputs rests on certain rules and characteristics present in the physical world and the way stimuli behave (Chun, 2003; Turk-Browne, 2012). Events and the behaviour of stimuli in the natural world rarely evolve completely randomly. Usually, the way events unfold and the way stimuli behave exhibit temporal regularities and relationships which can potentially be learnt and applied when interpreting variable visual information (de Lange, Heilbron, & Kok, 2018). One aspect of temporal structures and regularities that is very important to emphasise, is that such structures and regularities exist at varying levels of complexity (Barlow, 2001; de Lange, Heilbron, & Kok, 2018; Turk-Browne et al, 2009) which effects how they might be utilised by the brain.

1.8. Types of temporal regularity.

A common type of temporal regularity present in the behaviour of stimuli and indeed the wider environment is that the world is remarkably temporally stable over time or when events do occur they

are often repeats of one another (Fischer & Whitney, 2014; Liberman, Fischer, & Whitney, 2014; Liberman, Zhang, & Whitney, 2016). While we may not really think much about temporal stability as objects that are stable tend not to be behaviourally pressing, stability is nonetheless extremely prevalent. Examples of environmental stability are practically everywhere. If you look around your office or place of work right now, it is highly likely that very little is changing. Walls and windows remain in the same place, your desk does not suddenly appear in another part of the room and a book left on a shelf remains in the same location unless moved. Temporal stability can also occur in the behaviour of many stimuli in which the actions are simple repeats of themselves. The way people walk follows a similar pattern and the way a key goes into a lock and almost always turns clockwise provides past information that can be used to help minimize uncertainty. A commonly stable environment and repeated common behaviours means that in some instances a good estimate of a stimulus value of interest within a stream of noisy visual information is that next value will be the same or similar to the current values, at least over a short to medium time span, or that an action or event we have just observed will be repeated again.

In addition to an unchanging world or the simple repeats of behaviours a different type of temporal regularity are the conditional probabilities that exist between stimuli and events in the world (de Lange et al., 2018; Friston, 2010). For example, one simple type of conditional relationship are cues that signal a certain outcome. If you cook food in a microwave and the buzzer sounds we can make a judgment that as we have heard the buzzer our food is cooked. Importantly, more complex conditional relationships also exist within sequences of events that evolve over time. If you walk through a busy train station, you need to negotiate your way through lots of people many going in different directions heading to different exits to get onto your required train. Here, there are potentially many different paths people could take. One way to predict the position of other people and where they might be next might be by combining sequential information about how people will transition from the current to a future position based on the previous n-1, n-2, n-3, n-4… time points combined with knowledge of the exits and entrances of the station. By using conditional relationships provided by cues and sequences of events in our surroundings it is possible to decrease the level of uncertainty of sensory information which if judged on an independent basis might be subject to misinterpretation or errors (de Lange et al. 2018).

Empirical support for the idea that the visual system integrates information about the statistical regularities of stimuli and the environment is strongest in an area of research termed visual serial effects. Serial effects is an umbrella term study of a group of visual phenomena that indicate that perception that is not only driven by the stimulus information impinging on the retina at the current moment but by what has been observed in the recent past (Barlow, 2001; Gregory, 1970). Research areas that comes under serial effects include some very well studied areas such as visual masking, priming of pop out, serial dependence and adaptation to name but a few. Studies into such phenomena

provide some excellent examples for the general idea of temporal integration in visual perception and also highlight some limitations in the understanding of the exact type of strategies by which the visual system exploits such regularities

1.9 Broad evidence that the visual system integrates temporal structure when interpreting variable retinal measurements.

An area of serial effects research that provides a good illustration of temporal integration comes from visual masking. Visual masking refers to the phenomena that the current perception of a target stimulus is reduced by the presence of another stimuli called a mask (Breitmeyer & Ogmen, 2000). With respect to time three types of masking are usually tested; forward, backward, and simultaneous. In backward visual masking a target stimuli is presented for a short period of time and followed quickly by the "mask" (Kahneman, 1968). In suitable temporal conditions, the trailing mask can greatly reduce the perception of the target stimulus, even though the two visual events are separated in time (Breitmeyer & Ogmen, 2000; Kafaligonul, Breitmeyer, & Öğmen, 2015).The fact that the mask exerts an effect on the perception of the target even though the two events are distinct in time has been taken in support of the idea that the visual system retains a representation of the stimuli which it integrates in some way with the current retinal image (Breitmeyer & Ogmen, 2000). Because the two images are combined the perception of the current image is less accurate as it also contains information from the previous measurement (Breitmeyer & Ganz, 1976; Breitmeyer & Ogmen, 2000; Swift, 2013).

Further support for the notion that the visual system integrates information from the recent past into current perceptions comes from temporal form part integration. Temporal form-part integration refers to the finding in which two different sets of stimuli, presented at different times, are perceived by the visual system as a single integrated percept (Brockmole, Wang, & Irwin, 1985; Di Lollo, Hogben, & Dixon, 1994) Frequently, temporal form part integration studies use the two field paradigm in which stimuli usually comprise two stimuli showing, for instance, small dots or two parts of a matrix of dots with one piece or section missing in each stimuli that alone make no sense but together form a coherent image (Bachmann, 1997).The separated images are then presented successively with the time gap between images varied over trials. The task of the participant is to perform a judgment of what structure of the two incomplete patterns represent when combined i.e. letters. Importantly, the successful interpretation of a two a two field paradigm presentation can only be performed by integrating the first and second incomplete sections of dots into a complete figure. The required stimulus information to form a whole image cannot be drawn from either half image alone and can only come from the integration of both image sections. Usually, participants respond correctly at higher than chance levels when the half images are presented in sequences up to intervals of 120 ms between images (Swift, 2013). This repeatedly reported result provides support that the visual system

integrates information from images observed in separate time windows to help create a coherent image of two uncertain half or interrupted images.

An especially well studied phenomena studied under the umbrella of serial effects that illustrates the way past information can improve the effectiveness of behaviour is repetition priming. Repetition priming is a phenomena in which the behavioural response to a stimulus, usually measured in reaction times or accuracy, is improved by the repeated presentation of a stimulus (Kristjánsson, 2006; Yoshimoto et al., 2013). A number of priming of pop out tasks have provided strong and illustrative evidence for the contextualizing and performance enhancing input of previously observed stimuli. Commonly, in priming of pop out studies participants are asked to search for a stimulus of odd dimensions such as colour or shape relative to distractor stimuli of a similar but distinct nature that have another distinguishing feature such as a notch missing or an orientation marker (Becker, 2008; Goolsby & Suzuki, 2001; Maljkovic & Nakayama, 1994; Magnussen & Greenlee, 1999; Olivers & Meeter, 2008). Participants are then asked to state the nature of the distinguishing feature, i.e. is the notch to the right or the left or what orientation is the marker on the stimulus. Results normally report that if the target stimuli shares colour or shape with the previous target stimulus, even if the distinguishing feature that they are asked to report on is different, then reaction times are decreased or accuracy improved (Becker, 2008; Goolsby & Suzuki, 2001; Maljkovic & Nakayama, 1994) Crucially, because the priming observed is separate to the task demands then it might be interpreted as participants interpreting potentially uncertain information based on an integration of the salient features in stimuli.

Support from research into the various forms of serial effects strongly supports the idea that the visual system integrates information from the recent past into current perceptions. It is also highly likely that given the nature of such findings and our knowledge of the nature visual information received at the cortex that this is performed to aid the extraction and interpretation of uncertain task relevant stimulus information. However, there are a number of questions about such assumptions and findings due to the level of research serial effects operate at. All of the examples of serial effects literature we have presented operate at the empirical as opposed to the theoretical level. To elaborate on this distinction, the empirical level applies to observations about what happens under a certain set of experimental circumstances. On the other hand, the theoretical level attempts to explain findings from the empirical level and say how and why the findings occurred. For example, repetition priming gives an especially good illustration of the difference between the empirical and theoretical level research. The decrease in reaction times when stimuli are repeatedly presented is interesting but it is simply an observation. It is possible to speculate that under an assumption of a stable world that we might be tuned to expect what has been observed previously to be repeated and so be able to respond faster. This idea does not though say what happens to reaction times under more variable conditions or whether the reaction times might be due to the use of conditional relationships present in the sequences of stimuli as

opposed to simple repeats. Also, such empirical findings level studies do not touch on the underlying computations of integration or what the aim of integrating actually is. Is the aim of integration simply to increase signal to noise ratio or to actively reduce the amount of error in perception and responses which while appearing to be similar goals potentially arise from different computational strategies. Overall, on the basis of such research it could be valid to say that the visual system does integrate information over time but that it is also reasonable to say that such empirical research gives no theoretical explanation of the factors which mediate integration and the underlying computations that underpin integration or what the actual aim of integration might be. However, while much serial effects research does function at the empirical level there do exist ideas which do provide theoretical explanations of the how and why the brain integrates information over time.

1.10 Two predictive integration strategies: assume stability and predict based on the average of values observed over time or learn the conditional relationships between stimuli and predict.

Theoretical ideas about how and why the visual system integrates information about the statistical regularities of the environment have long concerned the thoughts of some of the most important figures in the history of cognitive science. These figures include including Helmholtz, Mach, Pearson, Craik, Attneave, Barlow and Gregory giving a clear indication of the level of such research. Due in no small part to the ideas of such crucial figures, an important idea has emerged within cognitive science. This idea, is that in order to help resolve uncertainty the brain makes forecasts or predictions about the content of its variable inputs based on past experience. Here though, it is important to raise an important point central to the current thesis. That is that the term prediction can be used in a number of ways and there is also more than one way to make a prediction. Over time a number of strategies about how the brain might predict the nature of its uncertain inputs have emerged and here we focus on two of them. Namely perceptual averaging (Corbett, Venuti, & Melcher, 2016) and perceptual inference derived mainly from the early works of Helmholtz (1867).

1.11 Perceptual averaging and perceptual inference.

Perceptual averaging is a simple way to make a prediction about variable measurements familiar to anyone who has worked with interpreting noisy signals that relies on a simple underlying assumption. This assumption is that there is some level of temporal stability in the stimulus under measurement but that measurements are also variable to some extent. Under this expectation a good way to predict a stimulus value is to base predictions on an average of values observed over time. Importantly, this type of prediction is while still a prediction more of a retrodictive type of prediction and relies on past information entirely. Perceptual inference on the other hand also involves a type of averaging to resolve uncertainty but in this case averaging also makes use of some of the more complex type of conditional sequential relationships present in the environment and is perhaps a more ‘true’ or prospective type of prediction. In perceptual inference, calculations can include basic regularities such as the repetitions or cues but can also factor in knowledge or beliefs about conditional relationships

present in more complex sequential information arising from the way stimuli behave or events occur in the world. Also, the time frames from which such learnt sequential and contextual knowledge can be applied to interpreting current sensory inputs can potentially be garnered from not only the previous seconds but potentially from much longer time frames. Crucially the ability to incorporate complex sequential relationships allows the construction of complex mental models that simulate future states of the environment based on how we expect the world to behave in the future in a more prospective manner (Friston, 2010). Examples which can help to distinguish the difference between perceptual averaging and perceptual inference can be observed in many day to day situations.

Anyone who travels to work using any type of large transit system such as an underground subway or rail network will have observed people making both predictions based on stability in events in the world and those based on more complex sequential or contextual information. Normally, on such networks the platform can be on the right or the left of the stopping train depending on the design of the station. People who wish to alight at their appropriate stop and miss the rush of people alighting from the same carriage must make a prediction about what side the platform will be on before they reach their stop so they can position themselves more effectively. This is where we can propose a hypothetical situation can help differentiate the two strategies. Say on one line of the network, the platform stays on the right hand side for five stops and then on the sixth stop switches to the left. At the sixth stop, people unfamiliar with the network, say tourists, will nearly always stand facing the right. This is an entirely viable strategy as the previous stops have all had platforms on the right. In the same situation it is also possible to observe a more inferential or model based type of prediction. People who are familiar with the line, say those who use it for travelling to work are able to model the sequential regularities of the platforms on the line. This means that when the train reaches the sixth stop they know to face the left side. This example shows the benefits of being able to utilise the sequential regularities of events in the world when making predictions. When events are uncertain basing predictions solely on what has occurred previously may be a good idea when things are stable but this type of prediction is in essence a prediction about the past and as we described will become inaccurate when events change in sequence or order. The ability to make predictions based on sequential regularities means that we can make more general and flexible predictions to help minimize uncertainty that are more about the ‘future’ which in the above subway case led to a more successful behavioural outcome. While the examples we discuss relate more to decision making type of predictions these general predictive strategies form the basis of two theories of visual processing. One which is largely an implied model we term fixed rate perceptual averaging (Kiyonaga, Scimeca, Bliss, & Whitney, 2017; Liberman et al., 2016) and a more formal model termed predictive coding (Gordon, Koenig-Robert, Tsuchiya, van Boxtel, & Hohwy, 2017; Huang & Rao, 2011; Spratling, 2008).

Fixed rate perceptual averaging is a term that can be applied to explain the finding that people's perceptions of current stimuli appears to revert towards the mean values of previously observed stimuli (Albrecht & Scholl, 2010; Albrecht, Scholl, & Chun, 2012; Jones & Dekker, 2018). Mean reversion is a century old finding first recorded in an experiment which showed that participants frequently choose a probe card that was too large when the cue card was small compared to the other cards presented in the experiment and selected a probe card that was too small when the cue card was larger (Hollingworth, 1910). Current findings indicting similar averaging behaviours come from both the spatial domain, in which the perception of a stimulus ensemble appears to represent the mean of shapes and sizes of objects in the current field of view (Campbell & Robson, 1968; Corbett, Wurnitsch, Schwartz, & Whitney, 2012a; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and in the temporal domain in perception appears to represent a reversion to the mean of stimulus values observed over time, a phenomena often termed serial dependence (Alais, Leung, & Van Der Burg, 2017 ; Cicchini et al., 2014.; Corbett, Fischer, & Whitney, 2011; Kiyonaga et al., 2017; Liberman, Fischer, & Whitney, 2014; Liberman, Zhang, & Whitney, 2016; Moors, Stein, Wagemans, & Ee, 2015).

Serial dependence is defined as bias in participant' judgments of a current stimulus value towards the mean of previous stimulus values (Fischer & Whitney, 2014). In serial dependence literature, mean reversion is interpreted under an internal assumption of a stable environment in which measurements at the retina are always uncertain to some extent (Fischer & Whitney, 2014). The general idea behind the function of serial dependence is that by representing perception as the fixed mean of values observed over time, variability in visual information from factors such as blinks and saccades are smoothed over allowing more accurate representations of the true values of stimuli of interest than those provided by a single uncertain measurement (Liberman et al., 2016; St. John-Saaltink et al., 2016). However, like any method of averaging or indeed statistical interpretation the implied fixed weighted method has its disadvantages and there are additional problems in regard to the way the fixed average account of temporal integration is somewhat less than clearly described in the literature.

There are two main issues with a fixed weighted averaging account of serial dependence. The first relates to the simplicity of the assumption of a stable world in which measurements are always uncertain. The second is a lack of an explicitly stated model for the fixed averaging perceptual integration strategy that is implied in serial dependence literature. The simplicity of applying a fixed average to all measurements is straightforward to understand. If you have a set of relatively stable statistical measurements from an experiment then averaging at a fixed rate would give a reasonable approximation of the underlying trends in the data. However, averaging at a fixed rate when values are highly variable but measurements clear leads to what might be considered unnecessary errors. This is because, as with any type of statistical averaging, some amount of past measurement values are always included in the current estimate. In visual terms, if the behaviour of stimuli is highly

variable then basing perceptions on a fixed average of past measurements means perceptions will not be indicative of the current state of signal as the perception is anchored on past measurements making them lagged in time to large changes in stimuli value. Also, the implied fixed average method of perceptual averaging suffers from a lack of an explicitly stated mathematical or computational model by which to guide experimental design and compare against other computational account of temporal integration. It appears that the general explanation of serial dependence is based on fixed weighted average models commonly used in signal processing but this is never actually stated. Without an explicitly stated computational account it is difficult to design experiments which test the fixed account of integration or compare against other accounts of temporal integration such as predictive coding which do have a more formalised if varied computational structure.

1.12. Predictive coding account of temporal integration: basic ideas

Predictive coding is a major theoretical movement within cognitive science and potentially represents a considerable paradigm shift in the way we think about vision (see Clark, 2013) for an overview of predictive coding theory). The general idea of some current and highly influential predictive coding models is that is that the visual system contains a series of hierarchical internal model(s) (Friston, 2010; Spratling, 2015). Each layer of the hierarchy contains an increasingly complex (from lower to higher) representation of the statistical regularities of the spatial structure and temporal regularities of the world. Based on the general parameters of the internal models the visual system constantly extrapolates or predicts the origin and cause of its expected neural and sensory activity such that superior hierarchal levels make predictions about activity at inferior levels via top down signals. Differences between predictions and measurements produce error or 'prediction error' signals which are sent back up the hierarchy to update the internal model and the subsequent predictions according to the nature of the prediction error.

In the literature, terms such as measurement and prediction can be used somewhat loosely so we define some important terminology. Here, we define prediction as an estimation of what the next sensory measurement will be. A measurement can be from the world as made by retinal ganglion cells as we have defined previously but in predictive coding also within early cortical regions (Friston, 2010). The next estimation can be next in time or in space. However, one very important aspect of predictive coding to be aware of is that predictive coding is essentially a computational motif. The term predictive coding simply means a neural process involving prediction and prediction error (Aitchison & Lengyel, 2017) and there are certain aspects of predictive coding that could be performed in multiple ways. Indeed, in the literature there exist a number of models that all differ in their application and frameworks (see Spratling, 2015 for a good account of a number of predicive coding models). An excellent example of predictive coding's general principles that provides an illustration of how spatial measurements, predictions and prediction errors are used to form

perceptions while at the same providing a ‘default model’ of predictive coding to highlight some of the differences between predictive coding models comes from Rao & Ballard (1999).

1.13. Rao and Ballard (1999) predictive coding in sensory cortex model.

Rao and Ballard’s (1999) computational framework provides a specific illustration of the nature of predictive coding’s hierarchical models and how the prediction, measurement, prediction error and update cycle function in the spatial prediction of image intensities. In Rao & Ballard’s (1999) hierarchal model, each level contains a representation of the spatial structures of an image intensities within an image. Each level in the hierarchy has an increase in receptive field size dealing with increasingly larger and complex areas of the image. In a three-level predictive coding model, level 0 will consist of a group of modules which deal directly with the measurement. Level 2 receives input from all the modules of Level 1 and at the same time feeds level 1 with prediction signals based on the probability of surrounding pixel intensities, while level 3 receives input from level two and at the same time predicts activity at level two. This hierarchical system functions throughout Rao & Ballards (1999) framework, with the highest level having the potential to receive input from all areas of the visual field and predict the whole image for the lower levels. Importantly, this cycle of prediction, measurement, comparison and prediction error happens constantly and with each iteration prediction error is reduced and predictions refined as more and more predictions and measurements are compared.

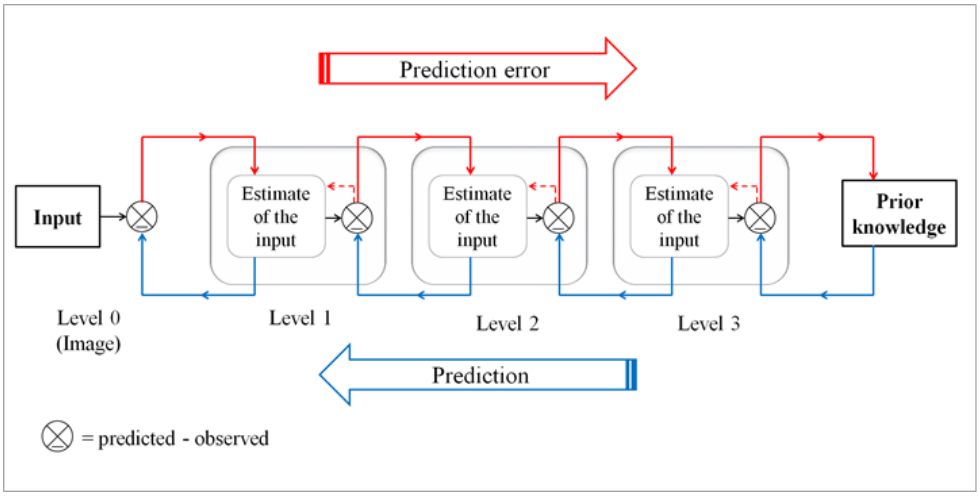


Figure 6. Schematic illustration of Rao & Ballard's (1999) predictive coding model.

Another key concept of predictive coding that is covered in Rao & Ballard (1999) is how to weight prediction errors, a term they call optimization function. For prediction errors to be a useful basis by which to inform and update perceptions they must recognize the variable nature of the natural world. For instance, greater confidence should be assigned to prediction errors in broad daylight than errors in prediction at nightfall (Hohwy, 2012). In Rao & Ballard (1999), prediction errors are weighted based on the inverse variance of the squared error between the current prediction and the mean of past measurements, such that a low sum of squares indicates a reliable prediction error while a larger sum of squares indicates less reliable prediction errors. With a higher level of reliability, internal models update and alter subsequent predictions by a larger amount relative to the size of the prediction error and with lower reliability the internal model updates and alters subsequent predictions by a smaller amount relative to the size of the prediction error. Importantly, the way prediction errors are weighted, varies between predictive coding models but normally involves the weighting of variances in prediction errors or measurements in some way and is adaptive to the setting in which predictions are made.

1.14. Missing pieces of the predictive coding puzzle: temporal predictive coding model and explicit perceptual averaging model.

One aspect to consider about predictive coding and a reason we use Rao & Ballards (1999) model to illustrate its general concepts is that there are a number of ways predictive coding could potentially be implemented. Rao & Ballard, (1999) is a computational model for simulating predictive coding in spatial image intensities only. The model is excellent for its purpose of explaining spatial predictive coding but like all computational predictive coding models it has its own take on the key aspects of predictive coding's general ideas. Within predictive coding literature, various other predictive coding models exist that all differ in application and configuration which may have led to some vagueness in

some areas of areas of predictive coding and perhaps contributed to some areas of visual processing including our area of interest, predictive coding in time, missing a specific account of predictive coding altogether.

In predictive coding literature, in addition to Rao & Ballard (1999), there exist models for predictive coding in the retina (Hoyosa & Meister, 2005), Spratling's PC/BC-DIM model (Spratling, 2015) again for spatial predictive coding and more general predictive coding models such as Fristons's free energy model (Friston, 2015) and linear predictive coding (Makhoul, 1975; O'Shaughnessy, 1988). Importantly, none of these models apply to the temporal domain directly. Also there are confusing accounts of the nature of the internal models probabilistic representation. The dominant predictive account is one of a hierarchical system based on conditional probabilities under Bayes optimal principles such that interpretation of activity at lower levels is conditioned by those above them (Clark, 2013; Friston, 2011; Friston et al., 2002). Rao & Ballards (1999) model also involves a hierarchical system seemingly implying the use of conditional probabilities but is primarily a linear model with no real need for complex conditional relationships and indeed never actually mentions conditional probabilities. Furthermore, exactly how prediction errors are weighted is often ignored in studies examining predictive coding in time. Again Rao & Ballard (1999), provide one way of weighting prediction errors, based on the inverse variance of errors but their method is based only on computational simulations of a model for spatial predictive coding. Overall, conflicting accounts of how predictive coding might be realised combined with a lack of a specific model for predictive coding in time highlight missing pieces of the predictive coding 'puzzle'.

The lack of a specific model for predictive coding is an important piece of the predictive coding puzzle that is currently missing. While we have discussed other models of predictive coding in the spatial domain and more general models this is not intended so much as a critique but simply to highlight the idea that there are potentially a number of ways predictive coding might be implemented depending on the sensory domain in question. There is no doubt that models such as Rao & Ballard (1999) and those from Spratling (2015) and Friston (2010) are excellent accounts of predictive coding but by its very definition the term 'prediction' applies to time. Indeed, it is the easiest type of prediction to conceptualise. When most people think about prediction, they consider events in time and the ability to predict what will happen in the future is absolutely vital to behaviour. Therefore, establishing the existence and understanding the computations underlying predictions in time must be central to our understanding of predictive coding. However, although predictive coding should apply to the temporal domain it is ambiguous as to whether previous studies that have examined predictive coding have actually observed predictive coding in time or whether findings are attributable to the more simple fixed rate perceptual averaging type of integration outlined in serial dependence which itself lacks an explicit model and been can be subject to alternate explanations.

1.15. Lack of specific computational temporal models make the behavioural signatures of fixed weighted perceptual averaging and predictive coding difficult to identify in current literature

Conceptually, predictive coding and fixed rate averaging appear markedly different. They each have contrasting aims, integrate different types of information and have different levels of complexity. However, due to a lack of clear and testable computational models by which to guide experimental designs it is open to question whether results interpreted under a fixed perceptual averaging in serial dependence literature have been correctly interpreted and also perhaps even more importantly whether predictive coding in time has been adequately distinguished (see table 1 below) . In order to illustrate the problems in experimental design caused by the lack of an explicit computational accounts of fixed perceptual averaging and the lack of an accepted model for predictive coding in time now we provide examples of current literature that highlights the nature of this problem.

Table 1. Predictive coding versus fixed rate perceptual averaging

Model signatures	Predictive coding	Fixed rate perceptual averaging
Fixed or blanket use of past information	No (Friston, 2018)	Yes (Fischer & Whitney, 2014)
Adaptive use of past and current information	Yes (Friston, 2018)	No (Fischer & Whitney, 2014)
Conditionally probabilistic internal models	Yes (Spratling, 2017)	No (Fischer & Whitney, 2014)
Reduction of prediction error over time	Yes (Friston, 2018)	No (Fischer & Whitney, 2014)
Passive estimate	No (Spratling, 2017)	Yes (Fischer & Whitney, 2014)
Active prediction	Yes (Friston, 2018)	No (Fischer & Whitney, 2014)

1.16. Fixed weighting of past and current measurements or not? Ambiguity in serial dependence research.

Perceptual averaging in the temporal domain is often termed serial dependence. Presently, serial dependence is an increasingly popular area of vision science which examines the nature of the observed bias or dependence of current perceptions on past stimulus values. Within an influential section of serial dependence research, it has been implied that serial dependence functions as a form of a fixed weighted average model in which the weights applied to measurements of past and current stimulus values included in perceptual estimates always remain at a fixed level (Fischer & Whitney, 2014; Liberman et al., 2014, 2016). If this idea is correct, a number of key identifying factors should exist in the current perceptual estimate of a stimulus value. One is that there should always be some influence of previous stimulus values in any current perception of a stimulus value. Another potentially identifiable hallmark, is that because the fixed account of serial dependence operates under a very simple assumption of a constantly stable but uncertain world the model does not include any concept of how the reliability in our perceptions of a stimulus or how the behaviour of a stimulus itself might affect integration and does not adapt perceptions to such situations. This means that the behaviour of a stimulus and how reliable the perception of a stimulus is does not affect serial dependence in any way. However, based on current serial dependence literature it appears that results from serial dependence studies may also be interpreted under at least the general principles of predictive coding which posits a more adaptive integration strategy.

Fischer & Whitney (2014), examined serial dependence in the perception of orientation. Their task involved presenting a series of orientated Gabor stimuli for 500 ms and then asking participants to estimate the orientation of the Gabors they has just seen by moving an adjustment bar (see figure 7 for a more detailed illustration) Interestingly, Fischer & Whitney (2014) claimed findings consistent with a fixed account of temporal integration and observed a bias in the judgment of both fully random and more stable counterbalanced Gabor stimulus orientations. One caveat with this interpretation is that the level of dependence while always present to some extent does appear to have been influenced by the variability of orientations and the level of variability in the measurement of the stimulus. When Gabor orientations had radically different orientations from previous orientations judgments serial dependence decreased and when orientations were more similar serial dependence increased indicating a role for stimulus variability. Another interesting aspect of Fischer & Whitney (2014) experiment is that while Gabor stimuli had a relatively high contrast (25% Michelson), the stimulus also had a relatively low spatial frequency (0.33 cycles per degree). The effect of low spatial frequency was to make Gabor orientations blurred. Blur has been shown to make judgments of stimuli more variable, quite feasibly due to an increase in measurement variability at the cortex (Kayargadde et al, , 1996). Furthermore, the use of a noise mask in between stimuli and judgment trials and presenting stimuli to the left or right of fixation and outside of the foveal representation may also have

contributed to increased measurement variability. These factors are important because if the role of measurement variability is a factor in temporal integration as in predictive coding (Friston, 2010; Gordon et al., 2017), it may be possible that the weighting attached to any prediction errors caused by the change in the Gabors orientation did not carry enough weight to update the new prediction to its full amount. If the prediction was not updated to the full amount of the prediction error then it may appear that the perception lies somewhere between the past and current stimulus measurement and appears serially dependent. However, in order to test the role of stimulus variability and measurement variability a more rigorous experimental manipulation that takes into account these factors is required.

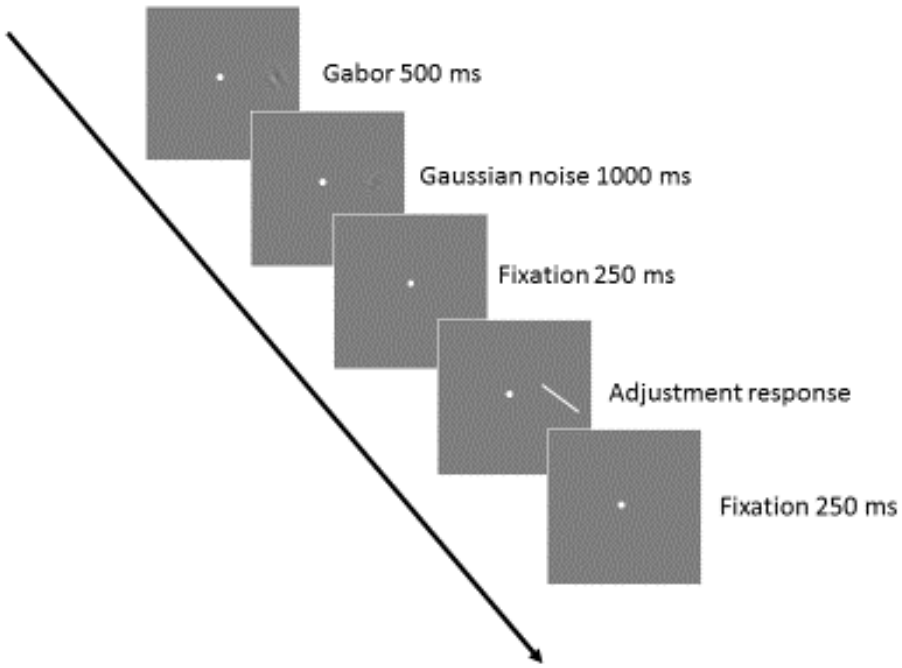


Figure 7. Fischer & Whitney's (2014) experimental design and task. Participants viewed a Gabor stimulus presented wither to the left or right of fixation interrupted by a noise mask and reported the perceived orientation of the Gabor by adjusting the orientation of a response bar.

In Manassi et al (2018), participants were asked to move a position marker to the position of a changing target stimuli. In a similar fashion to Fischer & Whitney (2014) participant's responses, in general, responses appeared to represent the average of the current and previous positions consistent with a fixed weighted average type integration. However, closer analysis of the data appears to support an alternative explanation consistent to some extent with predictive coding. Again, much like in Fischer & Whitney (2014) the level of serial dependence appeared to be modulated by both the similarity in position of the previous target stimuli and the presence of measurement variability which was caused by the use of an inter trial mask which is known to weaken the perceived contrast of the target (Breitmeyer & Ganz, 1976; Breitmeyer & Ogmen, 2000; Breitmeyer, Rudd, & Dunn, 1981).

Results showed that when positions of the target stimuli were more similar to previous targets serial dependence increased and decreased when further apart. Also, when the inter trial mask was removed no serial dependence was observed. Manassi et al (2018) acknowledges this result and explains it under the idea that the time reduction of the inter trial interval, reduced by the removal of the mask, was the causal factor but offers no theoretical explanation why this may be the case. A potential predictive coding explanation why serial dependence dropped out, is that when the mask was removed the level of measurement uncertainty in the weighting of any prediction error between the previous and current stimulus positions was reduced, meaning any error in the prediction of the current positions is weighted as being more reliable. With a more certain prediction error, it is possible that the current perception was corrected to the full extent of the prediction error produced by the change in stimulus positions. With a total correction to the new measurement no influence of previous stimulus values would be detected in the response and thus no serial dependence. Again though, such an interpretation is speculative and because the paper does not actually state that averaging is always fixed and does not provide any explicit model of the type of averaging implied it is hard to critique the findings in terms of the fixed weighted average theory.

A further interesting serial dependence study that could be subject to dual interpretations is Liberman, Zhang & Whitney (2016). This study examined the role that serial dependence has in interpreting partially occluded stimuli under the premise that integrating measurements over time helps to maintain a coherent perception when measurements are incomplete (Fischer & Whitney, 2014). Participants were tasked with judging the orientation of a partially occluded Gabor stimulus that was presented as either as continually moving or as a discontinuous series of orientations. Results again showed that serial dependence increased when orientations were more similar and decreased when further apart. Interestingly, and again providing contradictory evidence with a fixed weighted average account of serial dependence, in the discontinuous condition no serial dependence was observed. As no serial dependence was observed under more variable stimuli conditions this result could again be interpreted under a predictive coding or certainly adaptive accounts of integration in which the level of variability in measurements and stimulus behaviours play a role in updating predictions. Specifically, if a stimulus is highly variable then the new prediction should recognise this variability as a signal of change in the world and adapt to the new measurement value as quickly as possible. Once more though, while the role of stimulus variability appears to have played a role in the way measurements were integrated but without testing the modulating role of measurement noise in conjunction with stimulus variability in the way predictions update it may simply be the result of a bottom up driven process rather than a predictive coding type integration.

1.17. Predictive coding in time, or alternative explanations?

In a similar fashion to the way the fixed account of serial dependence research has been subject to a lack of clear theoretical guidance when designing adequate experimental designs, predictive coding in the temporal domain can also be considered to be subject to a similar problem. Within temporal predictive coding literature to date no concrete computational model has been tested and compared to the behaviour of human participants. Due to a lack of a clear and testable theory by which to base experiments upon it is open to question whether key aspects of predictive coding such the adaptive weighting of prediction errors have been tested fully in purely visual term. Furthermore, and perhaps most importantly, there are also questions about the type of probabilistic representations and complexity of information used to make predictions. To illustrate the limitations related to a lack of a clear theoretical model for predictive coding in time we provide evidence from a number of predictive coding studies that have claimed to observe predictive coding in time and debate the interpretation of such findings.

A neural phenomena that has been utilised to support the idea that predictive coding applies to the temporal domain is repetition suppression. Repetition suppression is defined as the diminished neural activation that results from the repeated presentation of a stimulus over time (Henson, 2003; Wiggs & Martin, 1998). Explanations for repetition suppression are keenly debated. The original explanation for repetition suppression was that reduced activity neural patterns could be explained by simple fatigue effects (Grill-Spector, Henson, & Martin, 2003) . By this account, the attenuation of a neural response is hypothesized to be due to an overall reduction in a neuron's firing rate as the neurons expend their energy or neurotransmitters. An alternative explanation of repetition suppression advanced by supporters of predictive coding, is that overall neural activity at lower levels of predictive coding's hierarchy is reduced by top down driven prediction signals when they are congruent with expected activity (Auksztulewicz & Friston, 2016; Grotheer, 2016a). Over time a number of studies have tested the predictive coding account of repetition suppression by manipulating the predictability of stimuli.

Summerfield Trittschuh, Monti, Mesulam, & Egner (2008) aimed to test the modulatory effect of prediction or 'expectations' on repetition suppression as indexed by blood-oxygen-level dependent (BOLD) imaging (higher BOLD signal is assumed less repetition suppression and lower BOLD assumed more repetition suppression). The study presented two types of trial. One type, called a repetition trial which comprised of pairs of identical faces and the second an alternation trial which showed different faces. To manipulate the expectation of which trial was most likely to occur Summerfield et al (2008) presented trials in two blocks termed high probability and low probability. In the high probability block the probability of a trial being a repetition trial was 75% and in low probability blocks the probability of being presented with repetition trial was 25%. Within blocks all

face images were trial unique which ensured that the probability of a repetition and not the repetition of a particular face was manipulated. As an incidental task was while observing trials, participants were asked to respond to occasional inverted faces the aim of which appears to have been to keep participants 'on trial'.

Summerfield Trittschuh, Monti, Mesulam, & Egner (2008) reported some interesting findings. Firstly, in high probability blocks BOLD signal was decreased with repeated trials in these blocks producing a decrease in BOLD signal of around 22% compared to alternation trials. In low probability blocks, repetition trials elicited only a reduction of 9% in BOLD signal compared to alternation trials. Statistical analysis showed a main effect of trial type on BOLD signal and more importantly a significant block type and stimulus interaction effect on BOLD signal meaning that the both the overall probability of seeing a repetition and actually seeing a repetition had an effect on BOLD signal. In terms of behavioural responses, reaction times to target inverted faces reported no significant difference between blocks. Summerfield et al (2008) concluded that the expectation of both stimulus repetition and block probability were both predictors of repetition suppression, with the context provided by block probability level superseding simple stimulus repetition. However, while results reported that the probability of a repetition and a repetition itself did reduce BOLD signal there are a number of questions related to the design and interpretation of the study with some aspects of Summerfield et al (2008) predictive coding interpretation of results open to debate.

One question arises from the type of 'predictions' in the behavioural aspect of the study. This is because participants were presented with a reliable cue. While it is of course still possible to make a prediction about the current input based on a cue but the use of cues can be thought of as providing a more simple associative relationship as opposed to a more 'forward' type of prediction about what will happen next based on sequential conditional probabilities. Perhaps some behavioural measure of what participants 'expected' was going to constitute the next face image should have been added to the study by inserting conditional relationships between stimuli into the sequences of trials and omitting trials and then asking for a prediction about the next trial might have been a better design. Another critique is that the study does not include any notion of the difference between the measurement of the stimuli in the world and the actual signals stimuli produce. In a number of influential predictive coding models, when the measurement of the stimulus is less certain the weight attached to prediction errors is downgraded (Friston, 2018; Spratling, 2015). Presenting clearly visible stimuli cannot, therefore, test this notion. In addition, due to the rather simple probabilities assigned to blocks and trials, it is open to question whether the neural activity changes observed between high and were due to the probabilities of observing repetition or alternation trails. A possible alternative explanation is that the visual system, as proposed by a fixed weighted average account of integration, models the average of the past and current stimulus measurements and during more changeable alternation and less probable blocks simply signals more changes in mean values when the stimulus

changes producing more neural activity. A final criticism that also provides a basis for a fixed weighted average integration interpretation of Summerfield et al (2008) results, is that even when stimulus repetition was rare repetition suppression was never totally absent. In a fixed average account some trace of past stimulus measurements always persists as the current estimate always contains at least some past stimulus history. The persistence of repetition suppression could reflect the persistence of past values as would occur with a fixed weighted account but in predictive coding this does not have to be the case. An attractive aspect of predictive coding is that given the correct conditions all past stimulus history can be discarded and the response will equate to the current stimulus which in theory should extinguish repetition suppression entirely. However, as the design of the experiment did not manipulate the reliability of sensory measurements (how well participants could see the stimulus) the experimental design was not adequate to test this aspect of the predictive coding.

An illustrative account from predictive coding literature that further illustrates the ambiguity in ascertaining which integration strategy has actually been performed comes from Summerfield & Koechlin (2008). In this study, the researchers aimed to manipulate predictions and predictions errors in upcoming stimuli orientations by providing them with specific perceptual templates that drew upon two classical psychophysical tasks. Task 1 was a two alternative forced choice task called the A/B task, because participants were shown two orientation lines of different colours (A, red & B, blue) on a grey circle separated by 60° on the same stimulus and then asked to report if the grating of a subsequently presented Gabor stimulus was the same as A or B . The second task was a yes/no type paradigm, in which they were only shown the A orientation and then asked to state whether the orientation of the next subsequently presented Gabor stimuli matched the A orientation or not (~A). Of total trials 50% of target Gabors in the A/B task were A and 50% B. Also in the A/~A task, 50% of the targets were A and 50% were ~A. The type of the target changed randomly from trial to trial in both tasks and trials were presented inter leaved blocks of A/B and A/ ~A.

Summerfield & Koechlin (2008) made separate hypotheses about the level of prediction error each trial type should produce. The first hypothesis was that highest level of prediction error should be observed in the non matched A/B trials and lowest in matched A/~A trials. In A/~A when the Gabor matched A prediction error should be zero but when orientation was a non match or ~A, prediction error should be increased in direct proportion to the difference in orientation which in non match trials was always 60° in either direction from A i.e. 60° of prediction error. The prediction error hypothesis in the A/B task was that prediction error should always be lower than A/~A non match tasks but higher than A/~A match tasks. This was because the participants had been presented with competing priors on which to base predictions so as a consequence the prediction should be in between the in A and B orientations separated which as they were always separated by 60° when presented with either an A or B target Gabor orientation prediction error should always be 30°.

Behavioural results for the study in the form of reaction times (which Summerfield & Koechlin (2008) indexed as a measure of prediction error) agreed with Summerfield & Koechlin's predictive coding (2008) hypotheses. Participants were faster in matched A/~A and slower in ~A trials. Reaction times for A/B tasks were also as hypothesized to be in middle of times for matched and non matched A/~A. Summerfield & Koechlin (2008) interpreted these results as largely due to the speeding effects of fulfilled predictions on behaviour. When predictions were 100% accurate as in matched A/~A trials then responses required less processing and were thus quicker. By the same measure, in non-matched A/~A trials, predictions were least accurate and required more processing as the prediction error was larger than non match matched A/~A and A/B trials and were thus slower. This interpretation applied by the study to behavioural data while at first glance plausible is subject to alternative interpretations not least due to questions about to the studies important concept of prediction.

Summerfield and Koechlin (2008) claim they manipulate predictability by presenting prior stimulus orientations and target Gabor stimuli orientations with the same stimuli and manipulate predictability by changing task demands. However, it is very much open to question whether they manipulated predictability at all. Stimuli were not structured in any predictable sequence or tasks designed in such a way that asked participants what was going to form the next observation. In both trials types and blocks the dominant feature was the A orientation. As the A orientation was always present in both sets of trial and in task instructions this raises questions whether participants actually needed to 'predict' the next orientation. An alternate explanation is that the differences in reaction times are simply due to an attentional resource constraint. That is in the A/~A participants had to keep in mind only one template while in the A/B participants had attend to two templates and attending to two templates simply takes more time (Bell, Wyatt, Bootzin, & Schwartz, 1996; Flaherty & Coren, 1974) To better test prediction error a design which also manipulated the predictability in the sequence of trials in a way that manipulated the conditional relationships between A/B and A/~A would be a better way of manipulating predictability more in keeping with current theoretical ideas of predictive coding.

1.18. Summary of chapter one and experimental questions.

The aim of chapter one was to introduce the notion of uncertainty in visual perception and current theory about how uncertainty in perception is resolved by the brain with an emphasis on some of the current theoretical limitations within the literature. To this end, we first discussed the importance of vision and how visual information is made uncertain by the presence of a number of sources of variability from both external and internal sources. We then drew attention to the apparent juxtaposition between the variable nature of visual information and then speed and clarity of our subjective visual experience. Next, we introduced ideas about of the use of past information about the temporal regularities present in the external world that the visual system could potentially capitalize

upon to help resolve its uncertain inputs with a detailed distinction about the different types of temporal regularities present in the world. We then provided empirical evidence from serial effects research that the visual system does integrate past information but explained that much of serial effects research operates at the findings level which while vital lacks theoretical understanding about why and how the visual system integrates past information. Next, we set out two broad theoretical strategies of how the visual system could reconcile uncertainty, namely perceptual averaging and perceptual inference and then discussed two current specific theoretical accounts of perceptual averaging and perceptual inference which form the basis for our experimental chapters.

The two current specific theoretical accounts of perceptual averaging and perceptual inference we introduce in section three were fixed rate perceptual averaging and predictive coding. Fixed rate perceptual averaging is a more implied model in areas of perceptual averaging study such as serial dependence while predictive coding is more concrete type of theory albeit with multiple versions depending on the sensory domain. The issues we highlight in the presented literature, in regard to both models, is that while appearing substantially different some of the key behavioural hallmarks of each model are hard to identify and distinguish. We then proposed this is in part due to a lack of computational accounts for both predictive coding in time and fixed rate perceptual averaging that could be used to guide experimental design. Therefore, chapters 2, 3 & 4 compare how information is weighted over time based on theoretical predictions made by model called a Kalman filter which is very similar in many ways to predictive coding and a fixed weighed average model which is identical to a fixed weighted account of perceptual averaging. In chapter 5, we test the idea of the visual perception as a dynamic system constantly making and updating predictions by assessing its ability to use conditional relationships in sequences of stimuli to improve behaviours. Results from all chapters are discussed in the final chapter, which constitutes our discussion and conclusions. We provide full explanations of our models and experiments in the relevant chapters but for now we provide a general idea and flow of how our experimental streams and chapters will operate to aid the reader.

1.19 Research questions and flow of experimental chapters

The literature presented for both fixed rate perceptual averaging and predictive coding raised a number of issues that could benefit from more theoretically structured experimental analysis. The three issues we highlight and further examine here are as follows; the adaptive weighting of prediction error versus fixed use of past and current information stimulus values in visual information, the adaptive reduction of prediction error over time and the extraction and use of conditional probabilities in enhancing behaviour and making predictions. To research these questions we structure our chapters as follows;

Chapter two. The Kalman filter and the fixed weighted average models.

Chapter two, provides background. Concepts, terminology and mathematical explanation of the Kalman filter and fixed weighted average models and the theoretical motivation for using both models in relation to comparing fixed rate perceptual averaging and predictive coding.

Chapter three. Testing the adaptive versus fixed use of past and current information.

The first aspect of predictive coding and fixed weight perceptual averaging we wish to test is the adaptive weighting of prediction errors versus fixed use of past and current stimulus values outlined in perceptual averaging. To allow this comparison, we design experiments based on the ideas of the Kalman filter in which the weighting of prediction error adapts to the level of stimulus and measurement variability. We use serial dependency as a means to assess the weighting strategy underlying participant's perceptual estimates and compare experimental predictions inspired by the Kalman filter and the fixed weighted average models about what should happen to the level of serial dependence in perceptual estimates under different levels of stimulus and measurement variability to participant's actual responses.

Chapter Four. Testing the adaptive correction of error over time.

Chapter three, examines the adaptive reduction of estimation error posited within predictive coding against the fixed reduction of estimation error outlined in perceptual averaging. To allow this comparison, we design an experiment based on the ideas of the Kalman filter in which the reduction of prediction error adapts to the level of stimulus and measurement variability. We use participants step responses as a means to assess how error in participants perceptual estimates is reduced over time and again compare experimental predictions inspired by the Kalman filter and the fixed weighted average models about how error in perceptual estimates should reduce under different experimental conditions with participant's actual responses

Chapter Five. Testing the extraction and use of sequential information in enhancing behaviour

In chapter five, we switch emphasis from looking at the adaptive aspects of predictive coding and examine the ability of the visual system to use conditional sequential transition information to make predictions. To examine such an ability, we test an experiment that manipulates sequential transition probabilities using a Markov chain. Here, we examine how participants learn and use these probabilities by analysing reaction times when presented with more and less predictable transitions and how they apply such information to make predictions in a forward manner.

Chapter 6. Discussion and conclusion

Chapter 6, presents our discussion and conclusion. Here, we recap the validity of our methods, discuss our main contributions to the literature and conclude whether our findings provide enough evidence to support the predictive coding account of temporal integration in the visual system.

Chapter 2. The Kalman filter and fixed weighted average models- Concepts, math and motivations.

Chapter two fulfils a key role in the current thesis. Its purpose is to set out the concepts and motivation behind our use of the Kalman Filter as a candidate model for predictive coding in time and fixed weighted average models as a model for fixed rate perceptual averaging. We also provide in depth math and equations for each model on pages 42-46 of the current chapter. Both the fixed weighted average and the Kalman filter models play key roles in experimental chapters 3 and 4, which largely focus on testing the adaptive nature of predictive coding. We begin with the simplest of our models, the fixed weighted average model.

2.1. Weighted average model: concepts, assumptions and example.

The model of temporal integration that has been implied in serial dependence literature (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Corbett et al., 2011; Fischer & Whitney, 2014; Kiyonaga et al., 2017) is the same as a fixed weighted average model (see equations 2-5, pages 42 and 43 for mathematical proof of this concept). A fixed weighted average model is a simple generic temporal smoothing model used to estimate stimulus values contained within variable measurements. This type of model, is perhaps the most commonly used model in signal processing and is used for estimating stimulus values in many applications such as global positioning systems (GPS) and radar (Lucas & Saccucci, 1990). The basic premise behind the use of fixed and more general weighted average models in signal processing, is that given a set of variable measurements, a good way reduce estimation error is to take an average value on a number of measurements recorded over time (Takahashi & Miwa, 1994; Zetterberg, 1958). Much like the stated role of averaging in serial dependence in perception, the model assumes that the underlying stimulus value is to some extent stable over time and that more random variability in measurements is due to more random noise unrelated to the activity of stimuli (Hua et al., 2017). The estimation process outlined in a fixed average models is dependent on a specific averaging calculation (see equations 2-5, pages 42-43).

A fixed weighted average model, is based on an averaging calculation in which each measurement of a stimulus value included in the calculation is assigned and multiplied by a fixed weight before summing to an average value. The number of measured stimulus values included in the averaging calculation includes the current amount and potentially any amount of previously recorded values. For example, in a fixed weighted average calculation with three values, the current measurement value might always carry a weight of 0.5 in the new estimate and n-1 value 0.3 and n-2 value, carrying 0.2 (weights must add to 1). Normally, some values in the calculation carry different weights according to when they were measured. Typically, the weights attached to measurements degrade as a function of time so that most recent values carry more weight and less recent values carry less weight but this is not always the case. Although the calculations involved in a fixed weighted average appear

remarkably simple they can be very effective if certain assumptions are met but like any temporal integration model fixed weighted average models also carry a number of disadvantages.

Assumptions, Advantages and disadvantages of the fixed weighted average model.

The assumptions for effective use of fixed weighted averages are simple. These are that the stimulus component within a measurement is relatively stable over time but the measurement of the stimulus also contains includes more random noise components (Rukhin, 2009). If this is the case then averaging acts to smooth over fluctuations caused by noise thus extracting a more accurate representation of the underlying but unknown stimulus value of interest in a very simple and easy to implement way. It is true that many situations meet such assumptions and due to just how often time series data exhibits such properties, fixed weighted average models and their variants are perhaps the most widely used of all temporal integration models. However, due to the simplicity of fixed weighted averages assumptions and its calculations they also carry a substantial number of disadvantages.

The main problem for fixed weighted average models are large or sudden changes in stimulus value. If there is a sudden fluctuation in stimulus value that is drastically different to previous values estimates based on a fixed weighted average will be inaccurate as they will be lagged towards the outlying past stimulus value potentially by large amounts. Another important disadvantage of fixed averaging calculations in terms of lagged responses is that if a value is invalid or missing then the model cannot perform its averaging calculations. This leads to a fixed minimum delay in the formation of a new estimate until a sufficient number of values are recorded. The lagged and non adaptive aspects of fixed averaging calculations are perhaps the most obvious criticisms of fixed weighted average models but there are also more subtle disadvantages in comparison to more complex temporal integration models caused by an inability to take into account information about the way stimulus behaviours evolve over time. In fixed average models, the assumption is that stimulus values under measurement will be noisy and stable. Stability is only one type of statistical regularity present in the way a stimulus can behave over time and the type that perhaps offers the least amount of information when trying to estimate values over time and other models do incorporate such models in a way that can be updated over time to build in more complex aspects of the data to help improve estimates. To illustrate the advantages and disadvantages of fixed weighted average models we provide an example of how a fixed weighted average models provide an estimate of voltage in an electrical substation which is a realistic example of an application for fixed weighted average models.

Weighted average: substation example.

A good way to illustrate the way a fixed weighted average model estimates stimulus values with different variances and noise levels within its measurements is by showing how a fixed weighted average makes estimates of sharply changing voltage values. In this example, we want to estimate the level of voltage emitting from a substation that takes in electricity from a power station and increases

the charge ready to send over long distances via the power grid. The normal output from the substation is 0.4 mega volts (MV) but recently, the level of voltage sent to the grid from the substation has appeared to sharply fluctuate and then stabilize at values between 0.1 and 0.9 MV, indicating that there may be an underlying issue with the substation. However, the measurements made by the measuring system also appear to be more variable than usual and may not be accurate and before shutting down the station for repair the operations team need to provide a more accurate level of the true level of voltage sent from the substation. To do this, they use a fixed weighted average to estimate the true underlying voltage values. The fixed weighted average they use comprises two values with a weight of 0.5 attached to each value which they apply to four test data sets which each comprise 30 data points 1 second apart. All of the data sets contains different levels of voltage fluctuations and noise. In this example, we also assume we know the true voltage value although in the real world we would not. The output of the four tests are illustrated below in figure 1 which we use to highlight the advantages and disadvantages of the fixed weighted average model.

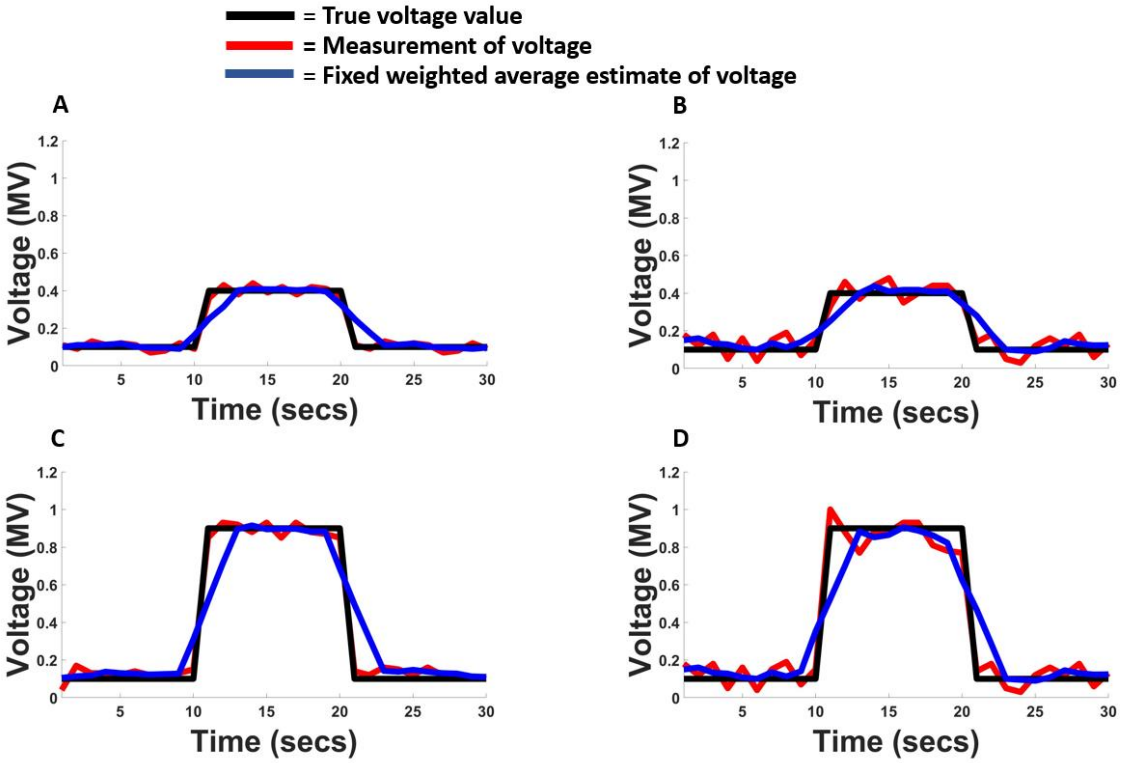


Figure 1. How a fixed weighted average makes estimates of fluctuating voltage values. Here we see the advantages and disadvantages of the way a fixed weighted average model makes estimates under different levels of stimulus and measurement variability. In sub figure A, the change in voltage is relatively small and the measurements relatively 'clean'. In this case, the variability in measurements from noise is smoothed over quite well and the estimates are more accurate than the measurements alone and while there is some lag during the change of the stimulus it is not too bad and overall the estimates are improved. In sub figure B there is a similar situation to figure A. In this situation, noise is higher and the measurements more variable. In this case, the fixed average model works very well to smooth out the noise and make estimates closer to the true value although as always with a fixed average there is still some lag but the estimates are improved. In sub figure C, we observe the major problems with averaging at a fixed value. Here, the voltage suddenly increases by a large amount and the estimates are much lagged behind the true stimulus value. In this case estimates are made worse than those given by the measurements. In sub figure D, we again observe the severe lag caused by the sudden change in stimulus. Again, noise is smoothed over but the cost to accuracy and is very high and this estimate would not be accurate or fast enough to be useful.

2.2 The Kalman Filter: aims, concepts and example.

A powerful approach for estimating the true value of uncertain stimulus values recorded over time is recursive Bayesian estimation theory (see Haug, 2012 for an excellent account of Bayesian estimation theory). A special case of recursive Bayesian estimation theory is the Kalman filter (Kalman & Bucy, 1963). The Kalman filter, is an algorithm that takes a series of variable measurements recorded over time and produces estimates of current and potentially future stimulus values that are more reliable than those based on individual measurements alone (Hu, Chen, Chen, & Liu, 2003; Kalman, Mi, & Bezier, 2008). Since its introduction in the 1960's, the Kalman filter has come to be regarded as one

of the most significant engineering results of the mid-20th century (Welch & Bishop, 2006). If the Kalman filters assumptions are met then the model is the optimal way to reduce the level of mean squared error in estimates of dynamic systems and is still a very useful estimator in an even larger class of systems which do not technically meet all of its assumptions such is its utility.

There are four main assumptions required for the optimal use of the Kalman filter. These assumptions are that a) the stimulus in questions behaviour is linear over time and follows a Gauss-Markov process b) noise in the behaviour of the stimulus is normally distributed, c) the observation process is linear and finally d) observation or measurement noise is also normally distributed (Grover Brown & Hwang, 1999; Praveena & Ravikumar, 2013). In the world of time series analysis and signal processing, many dynamic systems meet such assumptions and the Kalman filter has seen huge application in these areas. Applications include, lunar exploration and missile guidance, terrestrially based navigational tools such as GPS (Grover Brown & Hwang, 1999) and is also used in many computer vision systems for tasks such as feature tracking and stabilizing depth measurements (Chen, 2012; Nummiaro, Koller-Meier, & Van Gool, 2003).

The way Kalman filter produces a reliable estimate of both current and future values of stimuli is reliant on two recursive steps. We define these steps as the prediction step and the correction step. Importantly, the prediction step also uses a model of the behaviour of the stimulus in question which capitalizes on conditional transition relationships in the behaviour of the measured stimulus over time and rules of the situation the system operates in.

Step phases: aims, terms and concepts.

Before outlining the details of the prediction and estimation correction steps of the Kalman filter it is important to be clear about what each step is attempting to do and define some terms as the concepts and terminology in the steps be subtle and a little confusing. The overall aim of the Kalman filter, is as stated, to optimize the level of mean square error in estimates of stimulus values and each step in the Kalman filter plays a specific role in this overall aim. In the prediction step, the Kalman filter calculations take information about the current value of the stimulus based on a weighted estimate of the difference between the previous prediction and current measurement of the stimulus value made in previous estimation correction step (the Kalman filter is recursive and each of the steps can be considered to be ‘first’, although accounts generally start with the prediction step) and then factors in model based knowledge of the way the stimulus behaves over time and how accurate predictions are and calculates a new prediction about what the next estimated value might be (Kalman et al., 2008). Importantly, the Kalman filter recognizes that the behaviour of the stimulus may change and the accuracy of measurements can become variable, potentially making its model of the stimulus incorrect (Welch & Bishop, 2006). This means that the Kalman filter needs to update its estimates so that they continue to supply relevant information in order to keep its model of the stimulus in question

up to date and prevent predictions from diverging from their actual values and increasing mean square error (Zarchan & Musoff, 2000). This step is performed in the estimation correction step of the Kalman filter each time a new stimulus measurement is made and by combining both prediction and estimation update steps the model is able to recursively drive down error over multiple iterations. To distinguish between prediction and estimate, which admittedly can be unclear, we define prediction as the next expected value of the stimulus which is made in the prediction step and passed forward for correction in the estimation update step and estimation as the current estimate of the stimulus value.

The prediction step: make a prediction by combining our previous estimates and our stimulus model information.

The prediction step calculates two important values. The first value is the variability of predictions which gives an idea about how accurate predictions might be. The calculations for deriving prediction variance are perhaps the real 'trick' of the Kalman filter as they provide an indirect means to ascertain the behaviour of the true, unobtainable, stimulus in the world and help inform how to weight the estimates in the estimation correction step (see equation 12 for a general solution). Generally, if prediction variance is low then predictions are considered accurate and if prediction variance is high then predictions are considered unreliable. The second value, is an actual prediction about the next stimulus value based on the current estimated value made in the previous estimation update step and a model of how the stimulus behaves over time that incorporates prediction variance and some rules or knowledge about the behaviour of the stimulus and the environment it operates in.

The idea behind the use of a model in the Kalman filter, is that given some knowledge of the current estimated stimulus value and how its behaviour evolves over time, it is possible to combine both sources of information make a forward prediction about what the next stimulus value will be. However, while combining current measurements with model based information provides a good general basis for predicting the forward value of stimulus values there is a problem in relying on predictions and error variance calculations made in the prediction step alone for more than one or two forward projections. If the stimulus remains at a constant value then predictions remain accurate and prediction variance low as the stimulus is unchanging but when stimulus values change in a way not encompassed by the model or the level of noise fluctuates by a large amount then predictions will pick up error and diverge from the true stimulus value over multiple iterations, with divergence normally in proportion to the variability in stimulus values over time. To solve this problem, Kalman (1963) built in the estimation correction step of the model which feeds back new information in to the prediction step recursively every time a new measurement is recorded.

The estimation correction step: update the stimulus model, prediction variance and subsequent predictions based on new information.

The estimation correction step is crucial to the way the Kalman filter keeps 'on track' with both its stimulus model and new predictions as it allows the Kalman filter to update its model of the stimulus's behaviour. The update provided in the estimation correction step is based on a comparison of the previous prediction made in the prediction step and the current measurement of the stimulus values which when different produce errors termed 'prediction errors'. The magnitude of prediction error provides information about how much the stimulus might have changed since the previous prediction and how much the next estimate to be passed to the prediction step should update and also provides information about current prediction variability. Importantly, because all stimulus measurements contain varying amounts of noise and measurement error, prediction errors need to be weighted for reliability before updating the new estimate. The weighting of prediction errors in the estimation update step is modulated by an adaptive weighting factor called the Kalman gain.

The optimal Kalman gain has a closed form solution that depends on just two variance values: the variance of the current measurement of the stimulus (as opposed to the stimulus itself) which we term the proximal stimulus and proximal variance and the variance of previous predictions (see previous paragraph) which is largely derived from the variance in the behaviour of the external stimuli which we define as the distal stimulus and distal variance (see fig 2). The Kalman filter considers proximal variance as "noise" which it wants to filter out, while it considers distal variance to represent true changes that its wants to keep. Understanding how the ratio of distal and proximal variance affect Kalman gain and how Kalman gain modulates the way prediction errors are weighted and the recursive process of the Kalman filter is key to understanding how the Kalman filter works and indeed much of the current thesis.

How much to update our estimates: prediction, distal and proximal variance and the Kalman gain

Kalman gain is vital to understanding the use of the Kalman filter in the two subsequent experimental chapters. This is due to the way Kalman gain optimally modulates how past and current information are integrated and the rate by which error is reduced over time. Kalman gain is an adaptive weighting factor that runs from 0-1 and is calculated based on the ratio between prediction variance and proximal variance. When prediction variance is larger, which is normally due to changes in distal stimulus values, relative to proximal variance, a higher Kalman gain closer to one is produced. With a higher Kalman gain, the Kalman filter weights prediction errors as being more reliable and updates new estimates by a larger amount and corrects faster towards the full extent of the prediction error the closer to one the Kalman gain becomes i.e. a gain of one means the new estimate will update to the full extent of the prediction error instantly. Alternatively, when proximal variance is higher relative to prediction variance, normally due to factors that disrupt the measurement of the stimulus of interest,

then a Kalman gain closer to 0 is produced. With a Kalman gain closer to zero, prediction errors are considered less reliable and new estimates update slower and by a lesser amount relative to the prediction error i.e. a gain of zero means the new prediction totally ignores the prediction error and remains unchanged. Importantly, the whole update, measurement, prediction error and weighting process is adaptive and unlike the fixed weighted average model optimally updates its estimates to deal with more or less uncertain situations.

Advantages and disadvantages of the Kalman filter.

In a similar fashion to fixed weighted average models the Kalman filter has a number of advantages and disadvantages. One major advantage is that unlike fixed weighted average models the Kalman filter is adaptive to the level of stimulus change and measurement variability. When faced with a rapidly changing stimulus values (high distal variance) and low measurement variability (proximal variance) the Kalman filter can increase Kalman gain and change its estimates to match the level of change signalled by prediction errors entirely. The same idea applies to situations when measurements are highly variable and stimulus changes small, meaning distinguishing between noise and true change can be difficult. In this case the Kalman filter can 'hedge its bets' and reduce Kalman gain making estimates less responsive to untrusted measurement change and smooth over noise. Furthermore, due to the nature of the model which takes information from the adaptive estimation correction step, estimates can be updated to take into account previously unmodelled aspects of the data set and changes in the behaviour of the stimulus over time. This use of a predictive model also means that if a data point is missing the model can use the last prediction to some extent meaning the new estimate is not delayed. By factoring in and updating quantities such as prediction error alongside prediction, distal and proximal variance data the Kalman filter can also provide an adaptive weight that can further refine mean squared error almost instantly without the need to wait for additional values or attaching a fixed weight to previous values. Another benefit of the recursive structure of the Kalman filter is that there is no need for memory per se as all that is retained is the previous state estimate making the model computationally 'light. The final benefit of the Kalman filter we list and perhaps its main contribution to the field of signal processing is the way the model can determine the reliability of its predictions without direct access to the stimulus in question which when thinking about the applications of the Kalman filter in estimating remotely provided GPS or radar signals is imperative. However, despite the Kalman filter being a more adaptable and effective model given certain assumptions it does also have its disadvantages.

The main disadvantage of the Kalman filter largely stems, from its use of a model to help reduce estimation error which somewhat ironically is perhaps also its biggest strength. Using knowledge of the systems dynamics works well if the systems behaviour is known. Normally, in signalling processing applications this is the case. However, if the systems behaviour is not correctly modelled

then it can create errors in predictions which generally suffer from over or under shoots. This is because incorrect information is factored into the prediction step which skews predictions towards an incorrect assumption about the behaviour of the stimulus. Due to the recursive nature of the Kalman filter the stimulus model can be updated but this can still be a problem especially in the first few iterations before the Kalman filter has had a chance to correct its model.

Another drawback is that the Kalman filter has quite a lot of assumptions for its optimal use. It requires that measurements and noise be linear and Gaussian. These assumptions not always met in many dynamic systems and although there are variants of the Kalman filter that can deal with such instances the Kalman filter in its original form is not appropriate for dealing with these occasions. The concepts of the Kalman filter and integration processes we describe are more complex than that outlined in fixed average models and also factors such as proximal and distal variance are admittedly more abstract. Although we do provide full mathematical explanation later in the current chapter it is again useful to provide illustration of an example in which the Kalman filters principles can be outlined. For this purpose, we again make use of our voltage estimation example we used to illustrate the fixed weighted average model using exactly the same scenarios and data sets we used previously and also include an additional figure to explain distal and proximal variance.

The Kalman filter: Substation example

Before illustrating the way a Kalman filter would hypothetically estimate uncertain and fluctuating voltage values it is useful to provide an example of exactly what constitutes proximal and distal variance in our example.

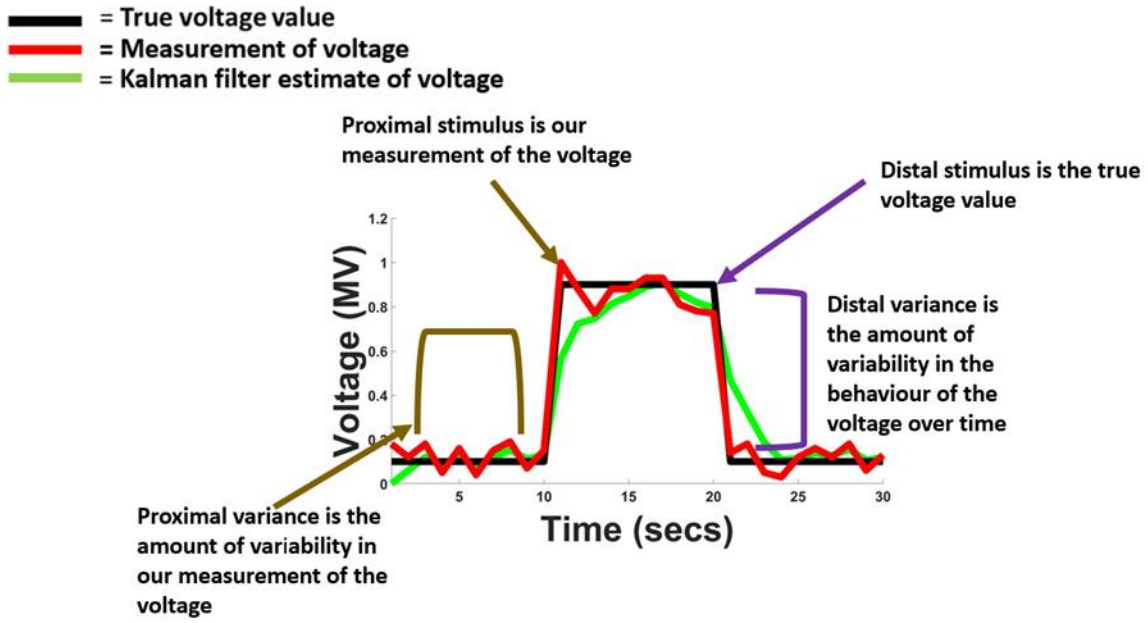


Figure 2. Example of the distinction between proximal and distal variance. Here, we illustrate the key distinction between proximal and distal variance. In our voltage example the distal stimulus is the actual voltage value while distal variance is the level of variability in the behaviour of voltages over time. Importantly, the value of the distal stimuli is unknown to the model and is what the Kalman filter is actually trying to model over multiple iterations until the error between the models estimates and true value becomes zero or 'steady state' is achieved. The proximal stimuli is the measurement of the voltage with proximal variance the variability in measurements over time. This is the aspect of the data that the model is trying to 'filter' as it represents unimportant variability that is unrelated to the behaviour of the stimulus and is considered noise.

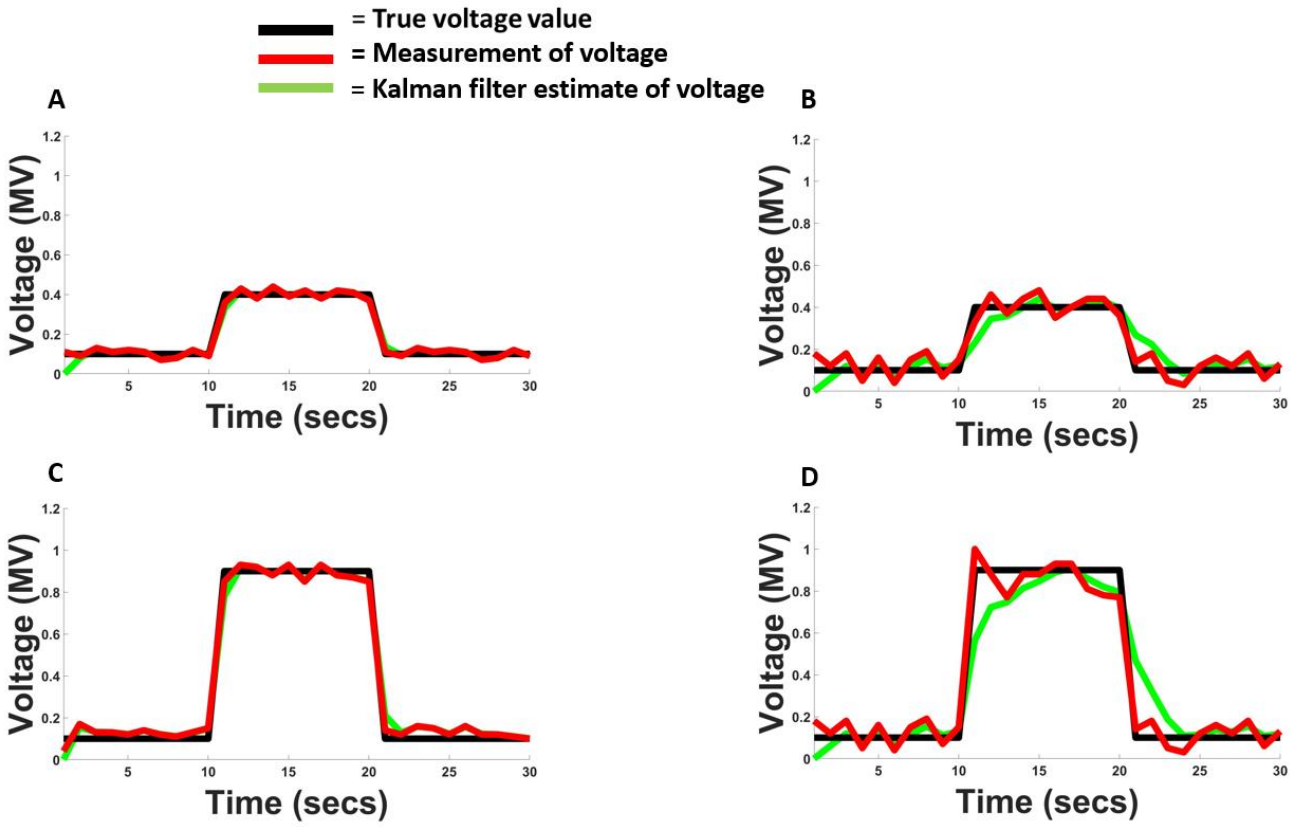


Figure 3. How a Kalman filter makes estimates of fluctuating voltage values. In this figure we observe the way the adaptive Kalman filter makes estimates under different levels of distal and proximal variability. In sub figure A we have a situation in which we have a relatively stable distal stimuli and low measurement variability. In this case, due to the low measurement variability or proximal variance, coupled with low prediction error variance due to a relatively stable stimulus we have a high Kalman gain close to one. In this case the estimate corrects almost entirely and instantly to the step in voltage values with very little error between the true stimulus value and the Kalman filters estimate. In sub figure B we observe a situation in which Kalman gain is reduced and is closer to zero. This is because the level of proximal variance is increased combined with low prediction error variance due to a quite stable stimulus. The effects of reduced Kalman gain is to smooth out noise and be less responsive to prediction errors that are trusted less. In sub figure C, we again observe the effects of higher Kalman gain we saw in sub figure A but they are more pronounced. The large step in voltage values causes a large increase in prediction error but because the level of proximal variance is low a gain very close to one is produced meaning prediction errors are corrected immediately despite the size of the change in value. Sub figure D again shows how the ratio of prediction variance and proximal variance modulate Kalman gain allowing the model to smooth over such variability and stabilize estimates. In this situation proximal variance is higher than in sub figure C and the model therefore reduces Kalman gain closer to zero which smooths out variability in measurements.

2.3 Equations.

Fixed weighted average and Kalman filter models

This section details the mathematical formulations of our fixed weighted average model and our Kalman filter model and a number of related issues (all equations for the Kalman filter are taken from (Orfanidis, 1988 with those for the fixed weighted average model based on generic averaging calculations). Both models deal with variables that change over time as well as variables that depend on different time periods. For simplification, we now outline notation. *X* will be used to refer to the true distal stimulus value, *Y* will be used to refer to measurements available to the observer (i.e. the proximal stimulus), and a hat identifies an estimate of a quantity. (E.g. $\hat{X}$ is the observer's estimate of the stimulus value). This section will adopt a notation that distinguishes between the current time point and the time from which previous information is being used. Subscripts will be used to identify different time points. A single subscript (e.g. Y_n) is used when the value only depends on a single time point. Two subscripts will be used to identify the time point and the information available to the variable. For example, $X_{n|n}$ means the value of X at time point n given all information up to *n*, $\hat{X}_{(n|n)}$ is the observer's estimate of X at time *n* given all information up to and including the observation at time n. While $\hat{X}_{(n|n-1)}$ is the observer's estimate of X at time *n* before having information, that is the observers prediction of what X will be at time *n*. For the following sections, we will assume the observers measurement Y reflects the true stimulus value X plus additive white Gaussian noise (EQ 1). Therefore, the reliability of the observer's measurement (σ^2_m).

$$Y_n = X_n + \varepsilon_m,\ \varepsilon_m \sim N(0, \sigma^2_m) \qquad \text{EQ 1}$$

2.3.1 Fixed weighted average model

$$\hat{X}_{n|n} = \sum_{i=0}^{p-1} W_{p-i}\, Y_{n-i} \qquad \text{EQ 2}$$

$$\sum_{i=0}^{p} W_i = 1 \qquad \text{EQ 3}$$

If the weighted average is computed using just the current and previous trial it simplifies

$$\hat{X}_{n|n} = (1 - W)Y_n + WY_{n-1} \qquad \text{EQ4}$$

Rearranging and grouping terms provides the linear equation

$$\hat{X}_{n|n} - Y_n = W(Y_{n-1} - Y_n) \qquad \text{EQ 5}$$

Equation 4 shows that for a simple weighted average of two past stimulus values the weight can be found by doing a linear regression that relates the current error to the change in stimulus values. As shown by equation 5 the slope of this linear regression (*W)* can be interpreted as the weight applied to the previous trial as shown in equation 5. Several previous studies have used this equation to estimate serial dependencies and therefore implicitly assume a weighted average of the current and previous trial (Fischer & Whitney, 2014; Manassi et al 2018; Liberman, Fischer & Whitney, 2014; John Saaltink, Kok & de Lange, 2016). Describing the problem in this form is equivalent to restricting solutions to be a finite impulse response (FIR) filter of the observations. The optimal weights (in terms of mean square error) can then be found by solving for the finite impulse response Wiener filter. However, while the finite impulse response constraint is useful in some contexts it is not needed for neural systems. While the Wiener filter can also be solved for without the FIR constraint, the Kalman filter provides a more useful conceptual framework for this problem. So we will now turn to explaining in detail the Kalman filter solution.

2.3.2. Kalman Filter.

The full form of the Kalman filter provides a general solution of estimating a changing variable that can be described by a linear system of equations in which measurements contain Gaussian noise. In this section we will use a restricted form of the Kalman filter that applies directly to our current experiments.

$$\hat{X}_{n|n} = \hat{X}_{n|n-1} + G(Y_n - \hat{X}_{n|n-1}) \qquad \text{EQ6}$$

$$G = \frac{\sigma_{pred}^2}{\sigma_m^2 + \sigma_{pred}^2} \qquad \text{EQ7}$$

In equation 6 the estimate of the stimulus value on trial n given all information up to trial n is $\hat{X}_{n|n.}$ The estimate of the stimulus value on trial n given all information up to the previous trial is $\hat{X}_{n|n-1}$ that is the prediction of the stimulus value. The current perceived measurement of the stimulus value is Y_n. The Kalman filter creates an estimate of the current stimulus values by taking the predicted value and recursively updating based on a proportion (G, Kalman gain) of the prediction error shown in equation 7.

Kalman gain.

The key quantity for the Kalman filter and conceptualising its adaptive integration strategy is Kalman gain. To calculate its value we need to determine the variance of the prediction (σ^2_{pred}). The solution for determining prediction variance is one of the key results of the Kalman filter. The prediction incorporates both past measurements as well as how the stimulus changes over time. Therefore, a model of how the distal stimulus changes over time is needed. It should be noted that the Kalman filter algorithm is an iterative algorithm and values are updated on every iteration based on the current measurement. The solutions presented below represent the values the Kalman filter will be after the estimate has achieved steady state. Furthermore, the general Kalman filter allows the model to be any linear system. In order to present the filter in a more intuitive way, here we will consider solutions to three restricted situations; uncorrelated stimuli, stable stimuli and a stimuli that follows a random walk. The uncorrelated model corresponds to the common experimental design choice to present stimuli in a random order. The stable model represents the extreme of a stable world in which the stimulus is stable and does not change. The random walk is used here and later in the thesis and provides a simple changing value that has some level of predictability. All three of these cases can be seen as special cases of a one term autoregressive model (EQ8&9).

$$X_n = c + aX_{n-1} + \epsilon_d \quad \text{EQ8}$$

$$\epsilon_d \sim N(0, \sigma_d) \quad \text{EQ9}$$

Performing Kalman filtering requires predicting the next measurement and then using the error between the measurement and prediction to update the current estimate of the stimulus by a proportional amount governed by the Kalman gain (G). Given the model in equations 8 and 9 the solution for these values are given by

$$\hat{X}_{n|n-1} = a\hat{X}_{n-1|n-1} + c \quad \text{EQ10}$$

$$\sigma_d^2 = \sigma_{pred}^2 - \frac{\sigma_{pred}^2 \sigma_m^2 a^2}{\sigma_{pred}^2 + \sigma_m^2} \quad \text{EQ 11}$$

Where σ^2_m is the variance of the measurement (proximal), σ^2_d is the variance of the stimulus and σ^2_{pred} is the variance of the prediction of the next value $\hat{X}_{n|n-1.}$ Equation 12, is a solution to the Riccati equation outlined in equation 11, solving the equation for the prediction variance.

$$\sigma_{pred}^2 = \frac{1}{2}(\pm\sqrt{2(a^2+1)\sigma_d^2\sigma_m^2 + (a^2-1)^2\sigma_m^4 + \sigma_d^4} + a^2\sigma_m^2 + \sigma_d^2 - \sigma_m^2) \quad \text{EQ12}$$

As we can see in equation 12, even in the case of a fairly simple stimulus model solving the variance of the prediction becomes quite a complex equation. In order to provide intuition for this equation, we will now show solutions to three specific cases.

Uncorrelated

For a presented stimulus that is randomly chosen, with no correlation to previous stimuli, prediction variance is equivalent to setting parameters *a* and *c* to 0.

$$X_n = \epsilon_d \quad \text{EQ13}$$

$$\hat{X}_{n|n-1} = 0 \quad \text{EQ14}$$

$$\sigma_{pred}^2 = \sigma_d^2 \quad \text{EQ15}$$

$$G = \frac{\sigma_d^2}{\sigma_m^2 + \sigma_d^2} \quad \text{EQ16}$$

This result represents the fact that if the stimulus has no predictability then the prediction should only represent the distribution the stimulus values are drawn. In this case, a uniform distribution. The gain is then identical to combining the measurement value with the mean of the stimulus distribution each weighted by their inverse variance.

Stable

To represent a stimulus that is completely stable we can set *a* and σ_d to 0.

$$X_n = c \quad \text{EQ17}$$

$$\hat{X}_{n|n-1} = \hat{X}_{n-1|n-1} \quad \text{EQ18}$$

$$\sigma_{pred}^2 = 0 \quad \text{EQ19}$$

$$G = 0 \quad \text{EQ20}$$

This finding of a Kalman gain of 0 may seem nonsensical as it means one should completely ignore the current measurement and purely respond only to previous values. However, these solutions represent the steady-state solution for the Kalman Gain. That is it's the gain value that is achieved by the filter in

the limit of an infinite number of measurements. What really happens is that if the world is completely stable, all information should be integrated with each new value getting progressively less weight in order to represent the sum total of all values. Because the value of the stimulus never changes, every measurement can be used to estimate the fixed value and with each new measurement the prediction becomes more reliable, and σ_{pred}^2 monotonically decreases to 0.

Gaussian Random Walk.

$$X_n = X_{n-1} + \epsilon_d \quad \text{EQ21}$$

$$\hat{X}_{n|n-1} = \hat{X}_{n-1|n-1} \quad \text{EQ22}$$

$$\sigma_{pred}^2 = \frac{\sigma_d^2 + \sqrt{\sigma_d^4 + 4\sigma_d^2\sigma_m^2}}{2} = \frac{\sigma_d^2 + \sigma_d\sqrt{\sigma_d^2 + 4\sigma_m^2}}{2} \quad \text{EQ23}$$

In this case, we have a stimulus value taking random steps drawn from a Gaussian distribution (EQ21 & 22). This creates a minimum level of predictability because the next value of the stimulus is related only to the previous time step. In order to build intuition, it is useful to consider the case where the measurement error is reduced to zero. In this case, the prediction variance becomes identical to the measurement and the variance is bound by the randomness of the stimulus and the Kalman gain becomes 1. This means that the prediction becomes identical to the measurement and the variance of the prediction is bound by the randomness of the stimulus (EQ23). As you add noise to the measurement the prediction variance becomes higher because a perfect estimate of the true value of the stimulus cannot be obtained. However this increase is sublinear.

2.5.3. Motivation for use of fixed weighted average models and link to experimental chapters 3 & 4.

Both the fixed weighted average and Kalman filter models play a central role in the next two experimental chapters. This is due to the marked similarities in the models to the method of integration implied in perceptual averaging literature in the case of the fixed weighted average model (Fischer & Whitney, 2014) and the more explicitly linked Kalman filter in the case of predictive coding (Rao & Ballard, 1999; Wolpert, 2007). Based on the concepts and mathematical formulations each model should produce behavioural aspects that can equate to aspects of fixed rate perceptual averaging and predictive coding that lack the experimental and computational clarity we outlined in chapter 1. Namely, the adaptive use of prior and current information and the adaptive reduction of error over time.

Use of fixed weighted average models and the Kalman filter to test the fixed versus adaptive weighting of past and current information to be tested in chapter 3

If the integration of past and current information over time functions at a fixed rate, as implied in serial dependence literature, then participant's behaviour should be approximated by a fixed weighted average model and certain signatures apparent. The most apparent signatures, would be that the level of past and current information in any estimate of a stimulus value should remain constant in any situation and at least some influence of past stimulus information should always be present (see equations 2-5). This is because the fixed weighted average model makes no distinction between variance caused by the change in stimulus or variance caused by viewing conditions and cannot adapt its estimates to recognize situations in which such factors may arise. Attaching a fixed weight to measurements also means that any amount of error produced by change in a stimulus value should remain constant when a change is repeated regardless of any viewing conditions or predictability of change. Another signature of a fixed weighted averaging strategy should be that responses should always be lagged or biased towards at least one previous stimulus value. Moreover, because the model does not contain any model of the behaviour of the external stimulus as proposed in predictive coding any predictive relationships in the sequential regularities in the behaviour of the stimulus would make no impact on the way estimates are formed.

If the integration of past and current information is adaptive, as proposed in predictive coding, then participants behaviour should be well approximated by the Kalman filter. The most obvious behavioural signature would be that due to way the Kalman filter adaptively weights prediction errors, the weight attached to past and current stimulus values should appear to change under different levels of proximal and distal variance. If we have a stimulus with high distal variance but low proximal variance, a higher Kalman gain should be calculated. With a higher Kalman gain, prediction errors are weighted as being more reliable and the new estimate is updated by a larger amount (see equations 13-20). Crucially, because when estimates update by a larger amount towards the new stimulus measurement value the estimate appears closer to the current distal stimulus value and further from the previous value (see equation 6). Alternatively, when the stimuli have low distal but high proximal variance a lower Kalman gain is produced. With a lower Kalman gain then prediction errors are weighted as being less reliable and the new estimate in the estimation correction phase of the Kalman filter will be closer to the previous prediction and appear biased towards previous stimulus values. Another behavioural signature should be the influence of sequential regularities or correlations in the behaviour of the distal stimuli over time (see equations 10-12). If the behaviour of the distal stimulus is correlated over time then this should produce some impact on the level of past information in estimates, especially in conditions of high proximal variance.

Use of fixed weighted average models and the Kalman filter to test adaptive reduction of error over time to be tested in chapter 4.

In the fixed weighted average strategy implied in serial dependence literature, the way estimates are formed are based on simple equations (see equations 2-5) in which a fixed weight is attached to each new stimulus value as they are measured. This means that this is a non adaptive model in which estimates do not update in a way modulated by any changes in proximal or distal variance. If this estimation strategy is approximated in humans then a similar pattern of lagged or speeded corrections in which the rate of correction is not modulated by the ratio of proximal and distal variance calculated as in the Kalman filter. In predictive coding, the way prediction error is reduced occurs in such a way that it reduces each time new comparisons between predictions and sensory inputs are over compared. This process can be considered an iterative process in a very similar way to the way the Kalman filter corrects error over time. If the way participant's correct error over time is approximated by the Kalman filter then we would expect past and current information and the way error reduces to be governed by the level of Kalman gain. With a high Kalman gain, less current information should be present in estimate and error should reduce more quickly and with a low Kalman gain more current information should be present in estimates which should reduce error more slowly. The next two experimental chapters tests these ideas directly.

Chapter 3. Testing the Adaptive Weighting of Prediction Errors: Serial Dependence and the Kalman filter.

3.1 Abstract.

Predictive coding offers a current theory of how the brain extracts and encodes behaviourally relevant information from the environment. However, while predictive coding is increasingly popular, certain tenets of the model lack behavioural support and computational understanding. One such idea is the precision weighting of prediction errors that mediate prediction updating. In predictive coding, precision relates to the best estimate of the reliability of prediction errors. If prediction errors are weighted by too high an amount predictions will update to change in the world that may not have occurred. Alternatively, if weighted by too low an amount predictions might be too slow to update to change. Technically, optimally estimating precision corresponds to optimizing the Kalman gain in the Kalman filter. In computational terms, Kalman gain underpins the optimal combining of different sensory information streams that differ in their variability as seen in multisensory studies but seldom tested in purely visual terms. Here, we model and behaviourally test the adaptive weighting of information over time outlined within the Kalman filter against the fixed use of past and current stimulus values implied in ideas of perceptual averaging using a fixed weighted average model. As it is not possible to isolate prediction errors directly via behavioural means we use serial dependence as a proxy measure of temporal integration and test model based hypotheses about what should happen to serial dependence under a number of experimental manipulations. According to the Kalman filter, the magnitude of serial dependence should adapt to different levels of measurement and stimulus variability while in the fixed weighted average model serial dependence should remain at a fixed level regardless of conditions. We tested both ideas using an experimental design from a study that had previously observed serial dependence (Fischer & Whitney, 2014) measurement and stimulus variability. In experiment one, we observed that when participants viewed highly variable randomly presented Gabors in two different contrast conditions (5 & 20%), less serial dependence was recorded consistent with our Kalman filter model predictions for a highly variable but clear stimulus. In experiment two, we reduced variability in our stimulus by introducing a correlated sequence into the orientations of stimuli across trials. In this experiment, we recorded a high level of serial dependence in the 5% contrast condition but no serial dependence in the 20%, again consistent with our Kalman filter model predictions for a less visible but more stable stimulus. Lastly, we report a supplemental experiment designed to replicate Fischer & Whitney's (2014) paradigm more closely. This experiment also reported no serial dependence, again consistent with our Kalman filter model predictions, providing general support for the predictive coding account of temporal integration.

3.2. Introduction.

Predictive coding is perhaps the most interesting theory of neural function to emerge in the last 30 years. Over time, a significant body of literature has emerged in support of predictive coding's ideas from many areas of psychology not least neuroimaging (den Ouden, Kok, & de Lange, 2012; Summerfield & Koechlin, 2008; Summerfield, Wyart, Johnen, & de Gardelle, 2011), psychophysics(Denison, Piazza, & Silver, 2011; Grotheer, 2016; Schmitt, Klingenhoefer, & Bremmer, 2018) and computational neuroscience (Friston, Parr, & Zeidman, 2016 ; Parr, Rees, & Friston, 2018; Spratling, 2008, 2015). However, despite a large body of literature in support of predictive coding and an increasing acceptance of at least its general principles, in some areas of sensory processing sections of predictive coding's computational 'puzzle' remain missing or incomplete. One such signature that lacks support is the precision weighting of prediction errors we discussed in chapter 1.

Determining the reliability of predictions relative to incoming sensory information is vital to predictive coding (see Friston, (2018) for an interesting debate about the importance of precison weighting in predicive coding). If the weight attached to sensory information is too low then prediction errors might not be weighted sufficiently. This could potentially lead to an over reliance on predictive information from the past and a delay in correcting to change in the world. Alternatively, if we attach too high a weight to potentially unreliable sensory input, we might over weight prediction errors leading to an over reliance on potentially unreliable sensory information, making perception too quick to correct errors when a real change may not have occurred. In predictive coding literature, there are a number of models (O'Shaughnessy, 1988; Rao, 1999; Friston, 2010; Spratling, 2015) which provide explanations of how prediction errors are weighted for reliability in different circumstances in a number of visual modalities. However, to date, there is no accepted tractable computational account of predictive coding that explains the way prediction errors are weighted in the temporal domain. One rich source of ideas that has great potential for providing models for testing adaptive weighting of prediction errors over time that have become prevalent outside of purely visual error processing is control theory.

Ideas from control theory: the Kalman filter and fixed weighted averaging models.

Control theory and the related field of control system engineering have vast experience in how to deal with uncertain inputs and correcting estimation errors in dynamic systems. One approach from control theory, increasingly used for understanding how errors in multi-sensory estimates should be weighted is recursive Bayesian estimation theory (see Haug, 2012 for an excellent account). An especially important concept contained in recursive Bayesian estimation theory, is the distinction between two related but distinct sources of variance. These sources of variance are variance in the behaviour of stimuli in the world or the distal stimuli and variability in the measurement of the distal stimulus at the measuring device or the proximal stimulus (Berger, 1985) that we explained in chapter 2.

Importantly, distal variance and proximal variance have different consequences for perception. Distal variance normally means that something in the world is changing that could be important so we want to keep this variability-on the other hand proximal variance tends to arise from factors such as poor lighting and weather conditions and can be thought of as noise which we want to remove. In Bayesian estimation theory, the interaction between proximal and distal variance has an important impact on the way predictions and estimation errors are weighted. A special case of recursive Bayesian estimation theory in which the relationship between distal and proximal variance on the way estimation errors are weighted is particularly well conceptualised is the Kalman filter (Kalman & Bucy, 1963) (See chapter two). The Kalman filter has been widely applied in an area of literature closely related to predictive coding termed Bayesian visou-motor integration that we have explained in depth in chapter 2.

The use of the Kalman filter to explain adaptive weighting of visou motor errors

Bayesian visuomotor integration deals with the problem of optimally combining noisy visual and motor information to produce an accurate estimate of our body in relation to its surroundings (Berniker & Kording, 2011) An interesting type of visuomotor experiment that often uses the Kalman filter as an explanatory model are reaching tasks (Baddeley, Ingram, & Miall, 2003; Knill, 2007; Wolpert & Flanagan, 2001). When reaching to point at a stimulus, estimates of hand position are uncertain due to noise in the motor system and in the visual estimates of hand and stimulus position (Knill, 2007; Kwon et al., 2015). This means the movement of the hand to the target is often inaccurate, especially initially. Visual feedback about the position of the hand can be used to guide the reaching but due to visual noise exactly how reliable this feedback is can be variable and sensory systems, if the aim is an optimal response, should represent this uncertainty in weighting visual feedback when correcting reaching errors (Denève, Duhamel, & Pouget, 2007). For example, when the hand is hidden is partially hidden or in the periphery of the visual field, visual feedback about the position of the hand should be weighted less than when fully visible or in the centre of the visual field (Knill & Pouget, 2004.). By the same measure, errors in hand motions should also be adaptively weighted. When the hand is moving quickly, motion signals should be considered less reliable then when the hand is moving more slowly (Wolpert, 2007).

Recent work has provided solid support that participants do indeed correct reaching errors in a way consistent with Kalman filter theory. A body psychophysical work has indicated a number of Kalman filter like response properties in the way humans correct reaching errors. A number of such studies have reported that humans use continuous on line feedback from the hand to correct errors in reaching which mirrors the adaptive process of the Kalman filter (Cluff, Crevecoeur, & Scott, 2015; Scheidt, Dingwell, & Mussa-Ivaldi, 2001; Wei & Körding, 2010; Wei et al., 2010). Furthermore, the weightings applied to both haptic and visual cues does appear to depend in part on the sensory noise (proximal variance) associated with each cue. (Baddeley et al., 2003; Harris & Wolpert, 1998)

Furthermore, evidence suggests that not only do participants weight the reliability of different cues based on their sensory noise levels but when variability is experimentally added to the visual feedback about the position of the hand (distal variance) this information is also optimally weighted when determining how much to correct previous errors (Burge, Ernst, & Banks, 2008; Burge, Girshick, & Banks, 2010; Cressman & Henriques, 2011; Saijo & Gomi, 2012) Based on such findings, which indicate that reaching errors are corrected based on the reliability of sensory information and level of stimulus change it is therefore logical to think that such an adaptive strategy may also be applied to purely visual integration . However, findings from visou-motor integration may represent the actions of distinct mechanisms than those of purely visual integration perhaps due to the high level of noise contained in motor commands (Lee, et al., 2016). Indeed, there is large amount of research about the way purely visual information is integrated which posits a simpler non adaptive strategy to interpret the world.

An alternate account: perceptual averaging and serial dependence

In the area of visual integration research termed perceptual averaging a much simpler way to interpret variable stimulus values has been proposed than that in predictive coding or the Kalman filter. Perceptual averaging studies have based their interpretation of temporal integration on the finding that the visual system appears to summarise the statistical values of stimulus values observed spatially and temporally (Ariely, 2001; Chong & Treisman, 2005; Cicchini et al., 2016; Corbett & Oriet, 2011; Corbett, Wurnitsch, Schwartz, & Whitney, 2012) as the mean of observed values. Such findings have been used to support the notion that because visual signals contain a substantial amount of noise but that the world is generally stable a good predictive strategy is simply to average together observed values (Liberman et al., 2014). In this way, variability from noise and retinal motion in individual retinal images is smoothed over making perceptions less noisy and more accurate than relying on individual potentially unreliable individual retinal samples (Corbett, Venuti, & Melcher, 2016). Although perceptual averaging is long standing finding (Gibson & Radner, 1937) it has recently received renewed popularity in the form of serial dependence.

Serial dependence is defined as the bias in current perceptions towards the average of previous stimulus values (Fischer & Whitney, 2014). In serial dependence literature the averaging strategy applied has been implied to function in a simple way. Specifically, it is implied that averaging is pervasive and functions at a fixed level across task demands or conditions (Corbett, Fischer, & Whitney, 2011; Fischer & Whitney, 2014; Liberman, Zhang, & Whitney, 2016). This is despite data indicating that serial dependency does appear to be adaptively modulated in many serial dependency experiments. For example, Fischer & Whitney report serial dependence to both totally random and counterbalanced stimulus orientations. However, when stimulus orientations were closer to one another they observed more serial dependence than when orientations were further apart with similar

findings indicating modulation of serial dependence by stimulus variability reported in a number of serial dependence papers (Bliss, Sun, & D'Esposito, 2017; Corbett, Fischer, & Whitney, 2011; Liberman et al., 2014, 2016; Taubert & Alais, 2016).

3.2.1. Theoretical motivation, aims and hypotheses of the current chapter.

The precision weighting of prediction error is central to predictive coding. The Kalman filter, already used in areas linked to predictive processing (Friston, 2018; Wolpert, 2007) offers an optimal model of how weighting might function. By providing the optimal Kalman gain, based on the ratio of distal to proximal variance, estimation errors are optimally weighted to prevent an over or under reliance on past and current information. However, most evidence for a Kalman filter type weighting comes from visou-motor work. It is possible that vision relies on a simpler averaging type mechanism which simply attaches a fixed weighted average to past and current stimulus values (Fischer & Whitney, 2014; Kiyonaga et al., 2017; Liberman et al., 2016). The current chapter aims to test the fixed weighting account of stimulus values outlined in serial dependency research versus the adaptive weighting of prediction errors as outlined in predictive coding and the Kalman filter. However, while the fixed weighting average integration strategy is relatively easy to test as it is based on actual stimulus values, isolating prediction errors is actually impossible behaviourally. This is because we cannot have access to the individual's actual prediction inside the brain if indeed this even exits. Therefore, we need an indirect experimental medium to test the adaptive weighting of prediction errors outlined in the Kalman filter and the fixed strategy outlined in fixed weighted average models.

In various parts of predictive coding literature a number of behavioural mediums have been linked to isolating and quantifying prediction errors. For example, reaction times (Summerfield & Egner, 2009; Summerfield & Koechlin, 2008) and task accuracy (Heeger, 2017). While we make use of both of these measures in subsequent chapters, at this initial stage we are more interested in testing the basic adaptive prediction error weighting strategy of the Kalman filter against the fixed weighting of implied in weighted average models as general commentary for the validity of predictive coding. A current and indeed useful medium for testing adaptive versus fixed weighting of information over time we have already introduced is serial dependence (Fischer & Whitney, 2014).

Serial dependencies are a good model phenomena for comparing the Kalman filter against the fixed weighted average model for multiple reasons. Firstly, considering that the Kalman filter has been previously linked with integrating visual and motor integration it is especially important to model a purely visual integrative phenomena. In this regard, a number of excellent control experiments by Fischer & Whitney (2014) have provided support that serial dependence is a visual phenomenon. Secondly, previous serial dependence studies have utilised some simple but effective experimental designs that lend well to manipulation under the ideas of distal and proximal variance (Fischer & Whitney, 2014) contained within the Kalman filter. Additionally, the way serial dependence is

calculated provides an easy to understand statistical measure of the way information is weighted/averaged by the visual system over time. Furthermore, both the Kalman filter and fixed weighted average models make strong and clearly defined experimental hypotheses about serial dependence in our experiments.

If the adaptive, predictive coding, account of prediction error weighting outlined in the Kalman filter is correct then Kalman gains and modelled trial stimulus weightings should predict the level of serial dependency over trials. With higher Kalman gains we should observe less serial dependence. This is because with a Kalman gain of one prediction errors ($Y_n - \hat{X}_{n|n-1}$) are weighted fully in a new estimate. When prediction errors are weighted fully estimates shift entirely towards the new Y_n stimulus measurement and past stimulus information has no influence making estimates appear entirely serially independent. Alternatively, with a Kalman gain of zero, total serial dependency should be observed. This is because with Kalman gains closer to zero prediction errors carry no weight and new estimates do not change and remain the same as the previous $\hat{X}_{n|n-1}$ estimate. Alternatively, if the fixed average account of stimulus weighting implied in perceptual averaging literature is correct then changes in Kalman gain should not predict the magnitude of serial dependence. In this case the magnitude of serial dependence should always be the same and remain at a fixed weight on trials regardless of any manipulation of distal or proximal variance.

3.3. Methods

Ethics

All calibration and experimental procedures were approved by the University of St Andrews Teaching and Research Ethics Committee. All participants gave informed consent.

Stimuli design and presentation.

In all experimental and calibration procedures visual stimuli were created in MATLAB 2015b (The Mathworks Inc) and presented using PsychToolbox (Brainard, 2007).

3.3.1 Proximal variance calibration experiment.

In order to provide an estimate of proximal variance that will allow the calculation of participant estimated Kalman gains we first undertake a proximal variance calibration experiment (see 3.8.2 page 64 for how we calculate proximal variance and relate it to our serial dependence experiments).

Participants

A total of eight participants undertook the proximal variance calibration experiment which followed a two alternative forced choice paradigm. (6 females, mean age 22, range 19-41). All participants were volunteers and recruited from the St Andrews SONA recruitment database.

Stimulus design and procedure

The proximal variance calibration experiment followed a two alternative forced choice paradigm in which participants were presented with an orientation discrimination task. Stimuli were oriented Gabor patches presented at either 5 or 20% contrast. All Gabors were embedded in Gaussian white noise (SD=15.5 cd/m^2) Noise following Gabors is also Gaussian white noise and covered the whole screen. All Gabor patches had a radius of 8 visual degrees and had a spatial frequency 0.5 cycles per visual degree. In trials pairs of Gabor stimuli were presented one after another in which the second Gabor was presented at a series of 7 different orientations from the first Gabor in the pair. These were +/- 1.5°, +/- 3.6°, +/- 5.7°, +/- 7.9°, +/- .10.1°, +/-12.16° and +/- 14.3. Each block presented 5 trials at both 5% and 20% contrast at each orientation difference giving 70 trials per block with each participant completing 3 blocks thus performing 210 trials in total. Fixations were positioned centrally (see Fig 1 below). The procedure of the experiment in outlined in figure 1 below.

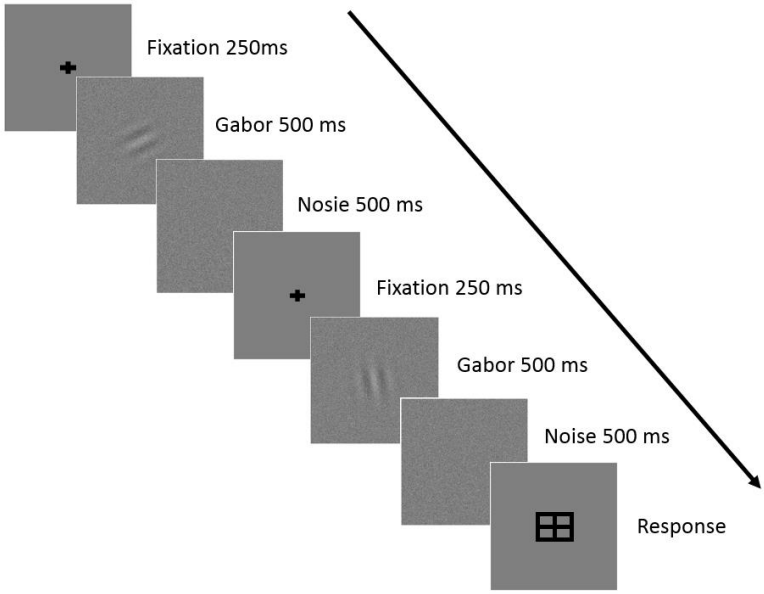


Figure 1. Proximal variance calibration experimental procedure. Participants were seated 57 cm from a CRT monitor. Each trial began with the presentation of a fixation cross in the centre of the screen for 250ms. Then a randomly oriented Gabor patch was presented for 500ms at 5% or 20% contrast depending on the condition then a noise mask for 500ms. Next a second Gabor was presented at one of 7 different angles ranging from (+/- 1.5°, +/- 3.6°, +/- 5.7°, +/- 7.9°, +/- 10.1°, +/-12.16° and +/- 14.3°) anti clock wise or clock wise from the first Gabor and then a second noise mask for 500ms. The task of the participant was to fixate on the fixation cross and then discriminate whether the orientation of the second Gabor was clock wise or anti clock wise of the first Gabor orientation. This was signalled by pressing j for clock wise and f for anti-clock wise. Each trial took approximately 6-8 seconds depending on how the response of the participants. After making a response, there was a 2-s delay during which only the fixation point was present before the onset of the next trial.

3.3.2 Main experiment one. Testing serial dependence under conditions of high distal variance versus low and high proximal variance.

Experiment one aimed to produce a high level of distal variance and a range of proximal variance conditions and to assess serial dependence under such conditions.

Participants.

Main experiment one had a total of 10 participants (eight females, mean age 23, range 19-41) Participants were a different set to those who had completed the proximal variance calibration experiment.

Stimulus design and procedure

To produce a high level of distal variance, Gabor orientations were presented in a fully random sequence between 0 & 360°. We manipulated proximal variance, based in part on our proximal variance calibration and presented Gabors in the 5% and 20% contrasts used in this calibration experiment but also presented Gabors at 10% contrast as an exploratory measure. However, upon analysing the basic error variances for the 10% contrast condition we did not detect any differences between the 10% and 20% contrast conditions so do include data for this condition in our analyses and modelling. Gabor patches had a radius of 8 visual degrees and had a spatial frequency 0.5 cycles per visual degree. All Gabors were embedded in Gaussian white noise (SD=15.5 cd/m2). Each participant completed 3 blocks of trials. Each block comprised 210 trials comprised of 70 trials in each contrast condition (5%, 10% & 20%). This meant each participant completed 210 trials in each contrast condition and 630 trials in total. Condition order presentation was randomized across participants. Blocks held the method of orientation change constant. The procedure and timings of the experiment are shown below in figure 2.

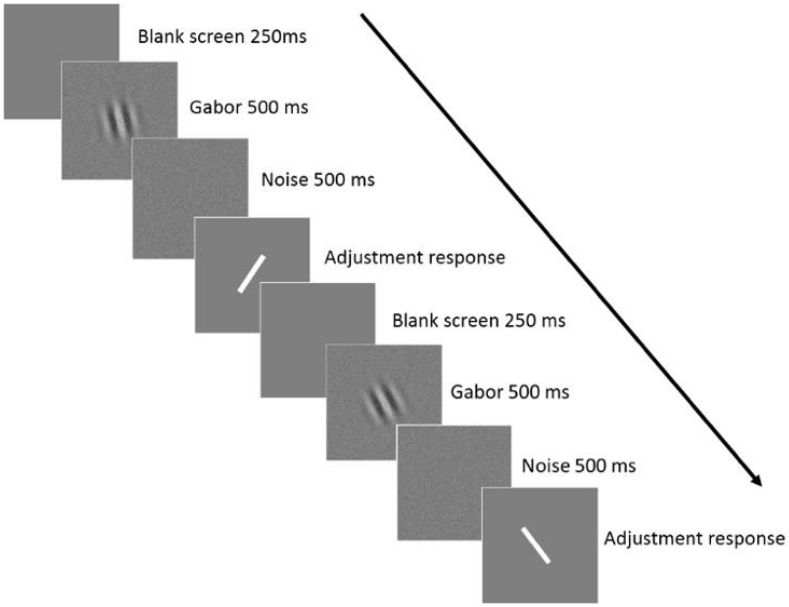


Figure 2. Stimulus design and procedure. Here we show how the experiment ran over two individual trials. Each trial began with the presentation of a blank screen (inter trial interval) for 250ms. Next a Gabor patch was shown in the centre of the screen for 500 ms, then a noise patch was presented for 500ms then an adjustment response bar was presented. The task of the participant was to move the adjustment response bar to try and match the orientation of the Gabor they had just observed. Each trial lasted approximately 6-8 seconds. After making a response, there was a 250 ms delay during which only a blank screen was present before the onset of the next trial.

3.3.2 Main experiment two. Testing serial dependence under low distal variance versus low and high proximal variance.

Experiment two, aimed to produce a lower level of distal variance than in experiment 1 and two levels (high and low) of proximal variance conditions and to assess the magnitude of serial dependence under such conditions.

Participants.

A total of 9 participants took part in experiment two (seven females, mean age 24, range 19-41). Participants were a different group to experiment one and the proximal variance calibration experiment. All participants were recruited from St Andrews SONA recruitment database.

Stimulus design and procedure

Stimuli design, timings and procedure were identical to experiment one with the only difference being the variability of Gabor orientations. To reduce the distal variance of our Gabor stimulus orientations were presented in a sequence that followed a Gaussian random walk (μ=0, SD=11.552), (see below for details and an explanation of the reason for this). As the 10% contrast condition recorded no differences in proximal variance to the 5% or 20% contrast conditions in experiment one, we did not include the 10% contrast condition in experiment two. Each participant completed 3 blocks. Each

block comprised 210 trials of 105 trials in each contrast condition. This meant each participant completed 315 trials in each contrast condition and 630 trials in total. Contrast condition order presentation was randomized across participants across blocks. All blocks kept the method of orientation change constant.

The reason we use a Gaussian random walk in our presentation of orientations in experiment two.

The term random walk, describes a stochastic process that follows a series of steps in some mathematical space. In a Gaussian random walk, steps are drawn from a normal distribution with the variability of values from the mean of values determined by the standard deviation of the Gaussian distribution. In terms understanding the use of a Gaussian random walk in the current experiment, which has a mean of zero, it simply means that the next orientation value can be considered to be the same as the previous with some level of variance, which here is a standard deviation of 11.552. This has the twin effect of introducing some level of predictability in the sequence of orientations and as how far from the mean orientations are presented is constrained by the standard deviation of the random walk built into the stimulus design code, the overall level of variability in orientations over time is reduced.

3.4 Analyses and statistical tests

Equations

Equations for both the Kalman filter for both a randomly changing and correlated stimulus which we use in the current chapter and a fixed weighted average model are fully outlined in chapter 2, pages 44-46.

Confidence intervals.

We report standard (parametric) 95 % confidence intervals of the mean.

Correction for potential ambiguity in responses using circular stimuli.

Recording responses and associated error from circular stimuli using a straight response bar can potentially be problematic. The reason for this problem is because the response line participants use to signal their estimate of orientation partially spans the diameter of the circle and points to two different angles. For example, a line at 60° also lies at 240° at the other side of the circle. This means that if an orientation on a trial is presented at 60° and the subject responds at say 65°, it is possible to record a response of both 65 giving 5° of error or 245° giving 180° of error. The way we and others (Fischer & Whitney, 2014) deal with this issue is to assume the response angle closest to the true trial orientation angle, as the angle the participant meant to signal and calculate the minimum angle difference between response and trial orientation as the response. This is performed by calculating the raw

difference between response and trial orientation and converting this to vector form using sin and cos and then converting this value back to an angle using the Matlab four-quadrant inverse tangent function (Mathworks Inc, 2012).

Proximal variance calibration calculations and relationship to Kalman filter equations.

To calculate proximal variance from our proximal calibration procedure we fit cumulative normal psychometric functions to the participants' responses data for both the 5% and 20% contrast conditions using the Palamedes toolbox's fitted variance parameter (Prins & Kingdom (2009). We take the variance ($degs^2$) of the width of the slope of fitted cumulative normal psychometric function as our measure of proximal variance i.e. a narrower slope indicates a less variable response (for example, a slope with a variance of 20 (deg^2) indicates a less variable perception of the stimulus in comparison to a slope with variance of 120 ($degs^2$) for the 5% contrast condition). To provide an estimate of proximal variance for a single stimulus presentation used in the main experimental conditions we divide the square root of the variance by two. Importantly, the variance we record represents the vital quantity of measurement variance (σ^2_m) or the reliability of the observer's measurement in the Kalman filter equations and can be used to form an estimate of participants Kalman gain for a response to a single stimulus presentation in conjunction with a figure for distal variance (see equation 7, page 48). To test for differences between proximal variances across participants for both contrast conditions we perform repeated measure t tests.

Estimated Kalman gains, Model fitted Kalman gains and fitted weights to participant data.

To calculate an estimated level of Kalman gain at the group level, we took our mean proximal variance measurements from our proximal variance calibration experiment and pre-set mean distal variance figures and followed the Kalman gain equation (see methods equations page 48). It should be noted, that due to the nature of presenting random orientations which produces a uniform distribution (high variance) and a Gaussian random walk (lower variance) the level of distal variance changes between each block and for each participant in each experimental session. This meant that the distal variance data was not the same for every participant. This meant that any statistical tests we ran on estimated Kalman gains for individual participants might not have provided unbiased comparisons so this aspect of our analysis is only intended to provide an approximate estimate of Kalman gains at a group level. In addition to estimating approximate Kalman gains, we computationally modelled participant Kalman gains.

Modelling of participant Kalman gains was carried out by running a nonlinear squares fit to ascertain which level of Kalman gain explained participant's responses (corrected for circular stimuli) most appropriately. A least squares model fit is a mathematical procedure for finding the best fitting regression slope or curve to a given set of data points. Here, this works by plotting a regression model based on the equations for estimation in the Kalman filter (chapter 2, page 43, and equation 6) at a

number of Kalman gain levels to participants' responses until the regression slope with the least residual squared error is found. The gain at which point is reached is taken as the participants' model fitted Kalman gain

Our modelling analysis of Kalman gain produces a specific level of Kalman gain in each subject and therefore is suitable for statistical comparison. Statistical analysis of model fitted Kalman gains between contrast conditions is carried out using paired sample t tests (Bonferroni-corrected). We report 95% confidence intervals. We also modelled participant responses with a fixed weighted average simulation model by running another nonlinear least squares fit to find the best fitting weights from the preceding 6 trial orientations to the current participant response.

Serial dependence calculations

To assess the magnitude of serial dependence across contrast conditions and experiments we calculated serial dependence based on Fischer & Whitney (2014). We first calculated and plotted the error in degrees on each individual trial on the Y axis (corrected for circular responses). Error is calculated as the participant response minus the actual stimulus orientation. Positive errors indicate responses clockwise of the true current trial orientation and negative errors indicating responses anti clockwise of the true trial orientation. Next, we plotted the relative orientation of the current stimulus in comparison to the previous stimulus orientation**.** Relative orientation is calculated as the previous (n-1) stimulus orientation minus the current stimulus orientation. To measure the magnitude of serial dependence over all trials we fit standard regression slopes corrected for symmetrical and circular stimuli to error and relative orientation. By this measure serial dependence is directly related to the steepness and intercept of the regression slope. In both experiments, statistical analysis of regression slopes used paired t tests (Bonferroni corrected).

N-back serial dependence analysis.

Previous studies have shown that participant's responses are not only serially dependent on the immediately preceding trial but that responses also depend on trials presented over the last 10-15 seconds (Fischer & Whitney, 2014). To assess how serial dependence changed as a function of time in our experiments we measured serial dependence between not just the current and immediately previous trial but between the current and previous six trials. This was calculated in the same way as serial dependence but instead of only comparing error on the current trial against the relative orientation of the current trial compared to immediately previous trial but also error on the current trial compared to the relative orientation of the previous six trials. Again, we fit regression slopes corrected for circular and symmetrical stimuli to participant's data and report standard deviations and 95% confidence intervals.

3.5. Results.

3.5.1 Proximal variance calibration.

Analysis of proximal variance data reported significant differences in the variability (degs^2) of cumulative psychometric functions fitted to participants responses between the 5% (M=141.66, CI [198.02, 91.88]) and 20% contrast conditions (M=38.50, CI [67.34, 19.05].) t (6) =4.01. p=.007. Significant differences between contrast conditions indicate that our experimental manipulation of contrast and our measurement paradigm were successful and this result is given extra validity by non overlapping 95% confidence intervals (see figure 4). One aspect of the results to draw attention to is how well participants could see the stimuli in the 20% contrast. Despite the presence of a noise mask, all participants were very accurate in observing changes between the first and second presented Gabors as soon as orientation changes became more than a few degrees (see figure 3 for an individual example).

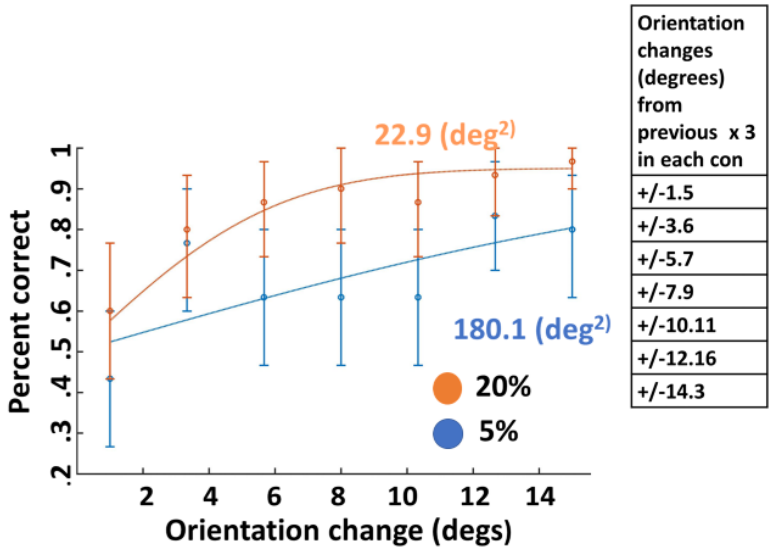


Figure 3. Typical individual 2AFC cumulative psychometric fit from our proximal variance calibration experiments All participants recorded less variability and greater accuracy in the judgment of the stimulus orientation relative to the previous in the more visible low proximal variance condition (20% contrast) as indicated by the steeper slope of the psychometric function in comparison to the less visible high proximal variance (5% contrast) condition which in all cases had a broader slope.

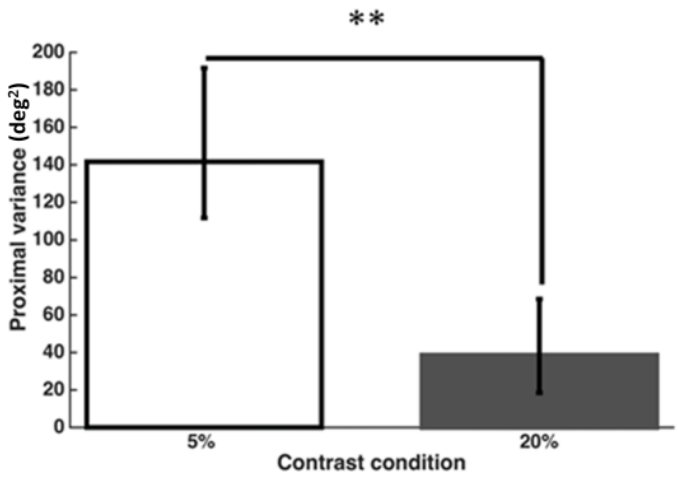


Figure 4.Group level differences in proximal variance obtained during our proximal variance calibration experiment-Note the large and significant difference in response variability between contrast conditions. In the 20% contrast condition participants performed well in nearly all cases. This contrasted with performance in the 5% contrast condition in which participants were less accurate in all cases.

3.5.2. Main experiment one.

Distal variance levels

The level of distal variance was determined by the variability of our stimulus orientations which in this case was high due to being presented completely randomly. As the level of variability inevitably changed over blocks due to the way orientations were presented randomly we needed to provide a mean figure we could plug into the equation for Kalman gain. Therefore, we took the mean variance figure from each block from all participants and divided it by the total number of blocks carried out in all experimental sessions to provide an approximate figure for distal variance. This figure was 2700 (deg^2).

Estimated Kalman gains

In experiment one we calculated estimated Kalman gains of 0.95 (5% contrast) and 0.98 (20% contrast) based on the calculations for the Kalman filter provided in chapter 2 page and our proximal variance calibration data for each contrast condition corrected for a single stimulus presentation (see figure 5 below).

Model fitted Kalman gains

Analysis of model fitted Kalman gain data mirrored results from our predicted Kalman gains and reported Kalman gains exactly the same as our estimated Kalman gains (see figure 5 below). In addition, no effects of contrast conditions on Kalman gain were observed in our analysis of model fitted Kalman gains with 5% (M= 0.95 SD=0.09, CI [1.00, 0.90]) and 20% (M=0.98, SD =0.07, CI [1.02, 0.93) t (9) =-2.270. p=0.059. No difference between conditions in modelled Kalman gains is confirmed by overlapping upper and lower bound 95% confidence intervals for the 5% contrast condition and 20% contrast condition, albeit with as expected slightly less variability in the more visible 20% contrast condition (see figure 5 below)

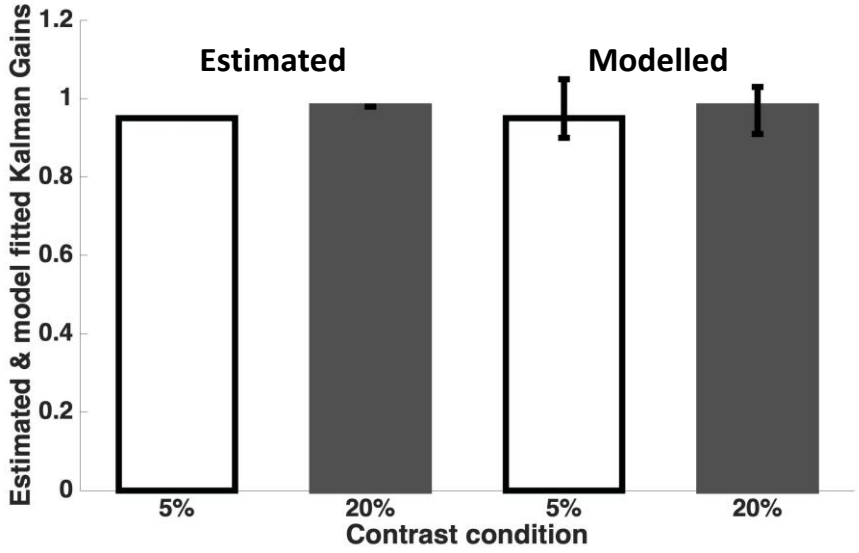


Figure 5. Estimated and model fitted Kalman gains for experiment one averaged across all participants. Note both predicted and model fitted Kalman gains are identical at close to one

Mean model fitted weights

In our fitted weight simulations in experiment one, we recorded weights at around one for the current trial orientation and around zero for both the 5% and 20% contrast conditions, (or in fact slightly negative weights in some cases), for all other 6 included n back trials in both contrast conditions (see figure 6 and table 1). This result, is consistent with Kalman gains of one we recorded in both contrast conditions in this experiment indicating that participant's responses, were all in cases centred on the current trial orientation. Interestingly, we also appear to observe a small of amount of negative weighting in both the 5% and 20 % contrast conditions. At this stage we cannot be sure of the reason for this and may be down to random effects or modelling errors and it will be interesting to see if it translates into serial dependence analysis directly.

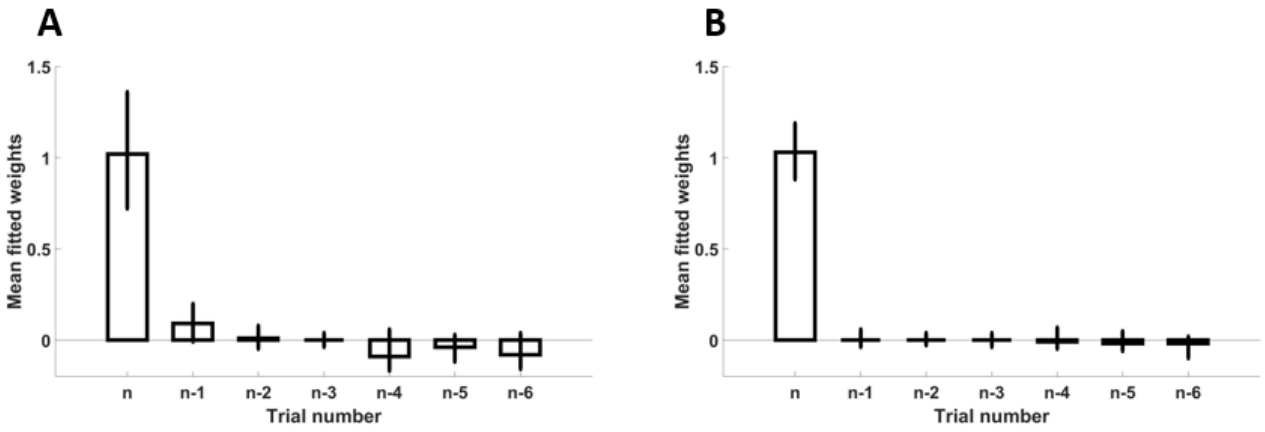


Figure 6. Mean model fitted weights for experiment one. Sub figures A & B illustrate mean model fitted weights in the 5% (A) & 20% (B) contrast condition on the current and previous 5 trial orientations. As expected based on our predicted and modelled Kalman gains we observed no indication of any influence of trial history in modelled weights or in our n back participant data. Model fitted weights are at 1 for the current trial orientation with no weight on any previous orientation. Note, as mentioned above we also appear to observe a small of amount of negative weighting in both the 5% and 20 % contrast conditions.

Table 1. Current and N back trial weights and 95% confidence intervals for experiment one

	5% Contrast		20% Contrast	
	Weight on trial	CI (LL, UL)	Weight on trial	CI(LL, UL)
Trial				
Current	1.02	[1.00, 1.36]	1.03	[0.84, 1.15]
N-1	0.09	[0.00, 0.20]	0.00	[-0.04, 0.06]
N-2	0.07	[-0.08, 0.08],	0.00	[0.03, 0.03]
N-3	0.00	[-0.05, 0.05]	0.00	[-0.04, -0.06]
N-4	-0.09	[-0.06, 0.13]	-0.01	[-0.06, 0.03]
N-5	-0.04	[-0.13, 0.10]	-0.01	[-0.02, 0.04]
N-6	-0.08	[0.09, 0.21]	-0.02	[-0.06, 0.03]

Serial dependence modelling predictions for experiment one.

Our modelling analyses allowed us to make very specific predictions about the level of serial dependence we should observe, not only between the current (n) and immediately previous trial orientations also between the current and previous five trial orientations (n-1, n-2, n-3, n-4, n-5 & n-6). Predicted and modelled Kalman gains of 0.95 for the 5% contrast condition and 0.98 for the 20% contrast condition make strong predictions that we should observe little if any serial dependence in experiment one for either contrast condition. Fitted weight analysis also provided a clear prediction that we should not observe any serial dependence towards any of the previous 6 trial orientations and that responses should be centred only on the current trial orientation in both the 5% and 20% contrast condition. Interestingly, model fitted weights also appear to support the idea of some form of negative weighting/serial dependency towards the previous trial orientations in the 5% contrast condition.

Individual serial dependence illustrations.

Before moving on to reporting our main analyses of serial dependence at a group level it is worth noting and illustrating that our model predictions were also accurate at the individual level. All participants (although some data was more noisy than others) recorded regression coefficients at or very close to zero indicating no bias towards previous values. This result, is entirely consistent with Kalman gains close to 1 and model fitted weights at 1 on the current trial orientation. Our individual plots also provide an idea of how well participants could perform at the task which although may seem simple is important in interpreting results. Based on the relatively low amount of error over trials, which was around the five to ten degree mark (shown by the width of shaded regions), it appears subjects could see the stimulus in both contrast conditions and perform the task very well as shown in figure 7 (below).

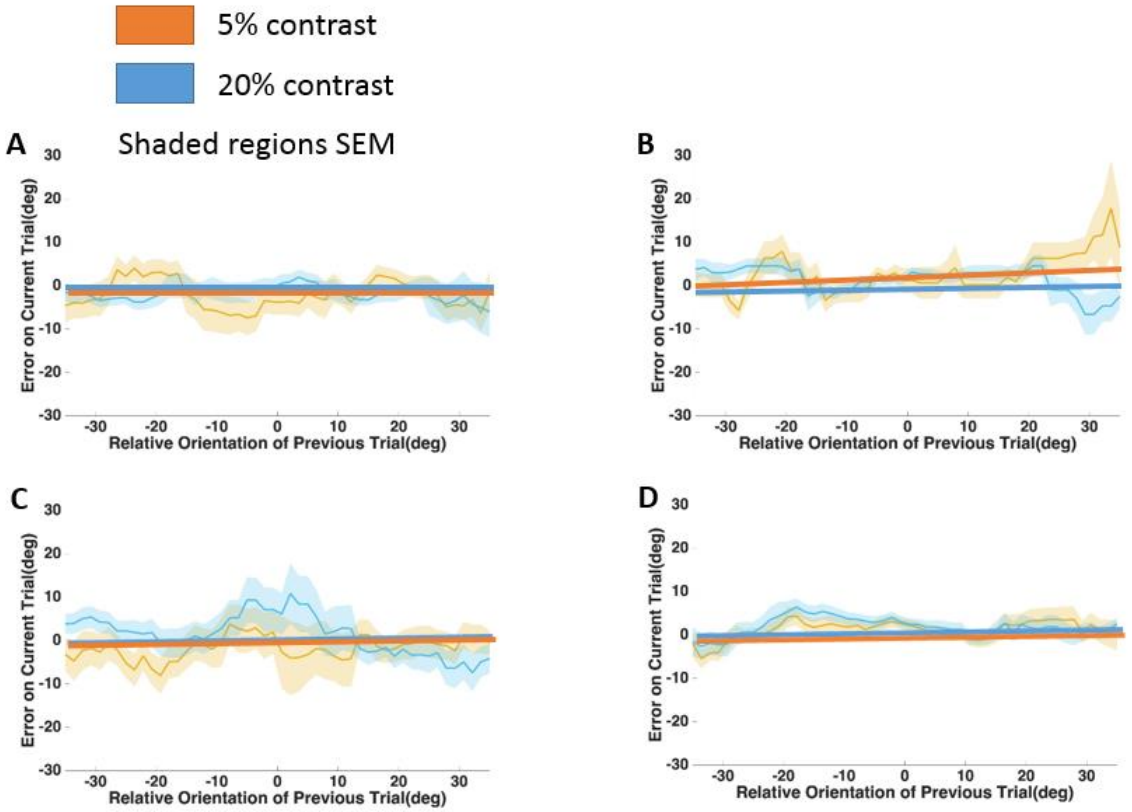


Figure 7. Four individual serial dependence plots for four individual participants. Positive values of the X axis indicate that the current trial was more clockwise than the current trial and positive errors on the current trial (Y axis) indicate that reported orientation was more clockwise than the true stimulus orientation. Regression slopes indicate the strength of relationship between error and the sined change in orientation. For example a slope of one would indicate that the subjects error to 100% captured by the change in orientation between trials and 100% serially dependent. By the same measure a slope of zero would indicate no relationship between error and the change in orientation and no serial dependence and in fact this is what we observe. Participants plotted in sub figures A, C and D reported regression coefficients at precisely zero while the participant plotted in in sub figure B may have exhibited some serial dependence (slope of 0.05) but at this low level could be simply attributable to statistical or response noise in an individual participant.

Serial dependency main analysis.

We hypothesised that the magnitude of serial dependence is modulated by the ratio of proximal and distal variance in our stimulus and commensurate to the level of Kalman gain. Analysis reported no differences in regression slope coefficients between contrast conditions and regression slope coefficients at or close to zero with 5% (M=0.03, SE =0.03, CI [0.00, 0.05]) and 20% (M=0.00, SE 0.02, CI [0.00, 0.02]), t(9)=2.530, p=0.320, This meant that while we did not record any differences in Kalman gains between conditions, we did observe a pattern of results indicating the level of serial dependence was commensurate to the level of Kalman gain. To clarify, the Kalman gains that we have obtained for experiment one are all close to one. This means that full weight should be attached to any prediction error. When full weight is attached to prediction errors, estimates should update fully and no past history (serial dependence) should be visible in estimates. Results strongly indicate participants responses are centred on the current trial (see figure 8) entirely consistent with this idea. However, such a result, while consistent with our model predictions is interesting, it is impossible to ascertain any adaptive effects of proximal or distal variance on serial dependence in this experiment as quite simply we do not observe any serial dependence. Also, this is clearly a failure to replicate Fischer & Whitney (2014).

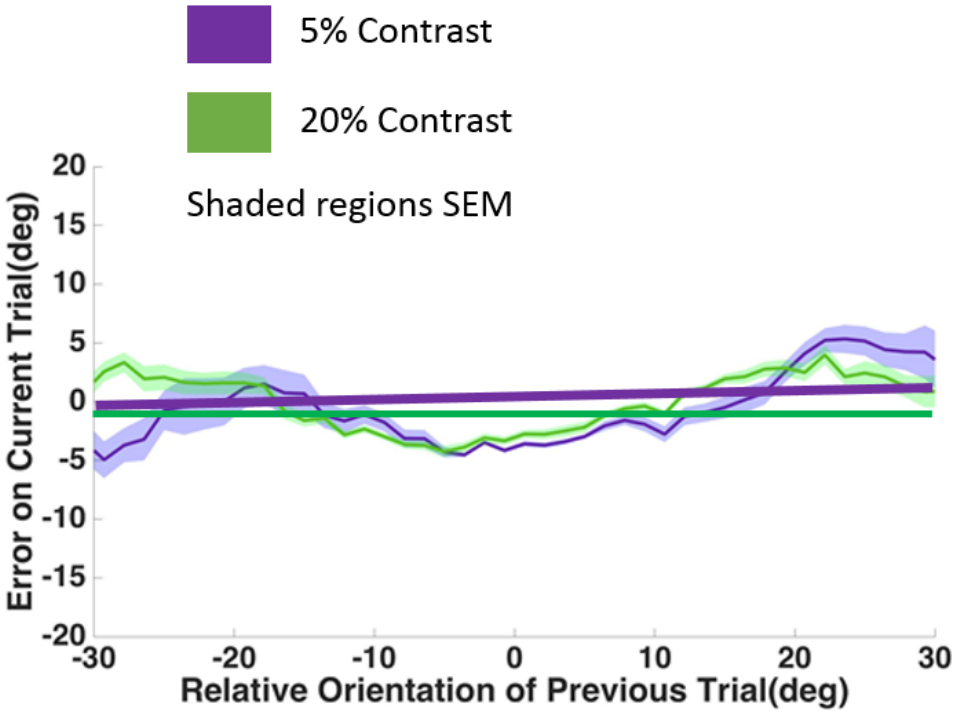


Figure 8. Group serial dependence plot. To recap, positive values of the X axis indicate that the current trial was more clockwise than the current trial and positive errors on the current trial (Y axis) indicate that reported orientation was more clockwise than the true stimulus orientation. Regression slopes indicate the strength of relationship between error and the sined change in orientation. Here we observe that subjects overall responded at the current trial orientation. While it may appear that there is some serial dependence between -10 & 10 we contend that the dip below zero on the anti-clockwise does not sufficiently rise above zero error on the clockwise (+) direction to be considered a 'serially dependent' distribution.

Experiment one n back analysis.

Results from our n back analysis were consistent with our main analysis of serial dependency and model predictions. Unsurprisingly, given we had recorded no serial dependence in our main analysis and based on Kalman gains of 1 and our model fitted weights at one for the current trial orientation we found no n back serial dependence for either the 5% or 20% contrast conditions (as shown below in figure 9 with exact figures provided in table 2). Interestingly, we also appear to observe some slight negative serial dependence or a repulsion effect in later n back trials. This may be attributable to chance due to limited trial numbers of perhaps more plausibly be related to adaptation after effects. While it may seem counter intuitive to associate adaptation with stimulus presentations lasting only 500 ms sub second negative after effects have been recorded previously Fritsche et al., 2017; Kanai & Verstraten, 2005) which can last for several seconds Fritsche et al., 2017). Fritsche et al. (2017). Again, though it should be noted that while our modelling predicted our results extremely accurately

it not possible to ascertain whether such a result supports our adaptive versus fixed hypothesis for serial dependence based on this experiment alone as none was observed.

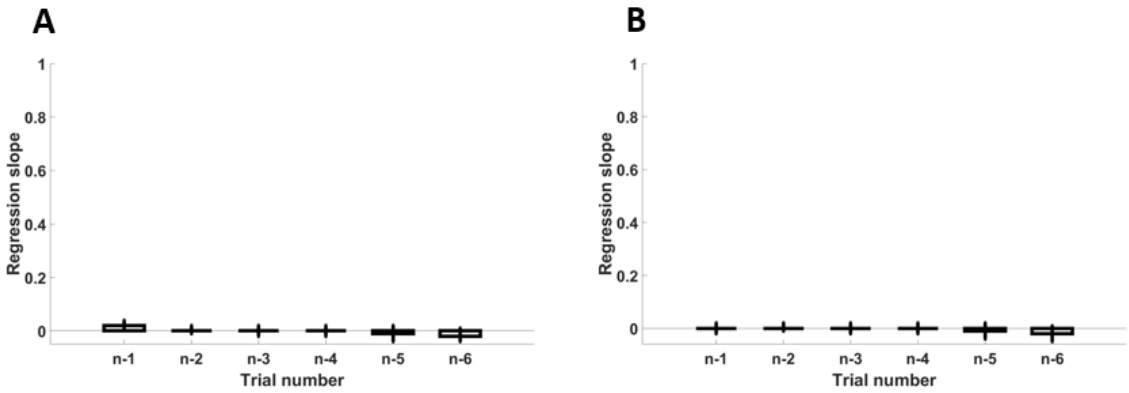


Figure 9. Mean n back regression slopes for experiment one. Sub figures A & B illustrate mean regression slope coefficients from all participants in experiment one in both the 5% contrast (A) and the 20% contrast (B). As we had observed no serial dependence in our main serial dependence analysis we were highly unlikely to observe any serial dependence in our n back analysis. This was confirmed with all regression slope coefficients being at or around zero in both the 5% and 20% contrast conditions.

Table 2. Serial dependence N back trial analysis for experiment one

	5% Contrast		20% Contrast	
	Regression slope	CI (LL, UL)	Regression slope	CI (LL,UL)
Trial				
N-1	0.00	[-0.01, 0.01]	0.00	[-0.01, 0.01]
N-2	0.00	[0.01, 0.01]	0.00	[0.01, 0.01]
N-3	-0.01	[-0.00, 0.02]	0.00	[0.01, 0.02]
N-4	0.00	[-0.01, 0.02]	0.00	[-0.01. 0.03]
N-5	0.02	[-0.03, 0.03]	-0.02	[-0.04, 0.02]
N-6	-0.03	[-0.04, 0.01]	-0.04	[-0.06,0.02]

3.5.3. Main experiment two

Testing serial dependence under reduced distal variance versus low and high proximal variance.

Distal variance levels.

To calculate the mean variance of orientations across all experiments we recorded the level of variability in stimulus orientations in each block and divided it by the number of blocks carried out in all experiments to provide an approximate figure for distal variance. This figure was 133 (deg^2) which is much reduced from the figure of 2700 (deg^2) we recorded in experiment one.

Estimated Kalman gains

Based on the calculations for prediction variance in a stimulus that follows a Gaussian random walk sequence (μ =0, σ =11.552) and our figures from our proximal variance calibration we calculated estimated Kalman gains of 0.62 for the 5% contrast condition and 0.83 for the 20% contrast condition. The interesting aspect of our estimated Kalman gains is that for both contrast conditions, Kalman gain is reduced away from one as we observed in experiment one. Such a result, also highlights the difficulty in reducing Kalman gain away from one. We have had to drastically reduce distal variance and make our stimulus extremely difficult to see but even here Kalman gains are still closer to one than they are to zero. Nonetheless, the fact that we have observed a difference between Kalman gains between contrast conditions is the important factor here and supports an adaptive weighting strategy.

Mean model fitted Kalman gains.

As in our estimated Kalman gains, in model fitted Kalman gains we recorded reduced Kalman gains from those recorded in experiment one but on this occasion only in one contrast condition (5%). We calculated a Kalman gain of 0.74 for the 5% contrast condition but a Kalman gain of 1.02 for the 20% contrast condition. Importantly, t tests reported significant differences in modelled Kalman gains between contrast conditions with 5% (M=0.74 SD=0.20, CI [0.88, 0.63]) and 20% (M=1.02 SD=0.53, % CI [0.99, 1.05]) t (8) =-3.629. p=.007. Note here, we have a discrepancy in model fitted Kalman gains and estimated Kalman gains. Model fitted Kalman gains have not reduced away from one as much as estimated Kalman gains. Results from model fitted Kalman gains again provide an insight into just how good visual perception is. Even in the 20% contrast condition, our stimulus was quite noisy and used noise masks, yet participants were able to respond almost entirely to the current trial orientation. In the 5% contrast condition, which it must be said was a tough stimulus to see, participants are still quite accurate given the noise level of the stimulus. Once again though, the key factor here is that we have a difference between Kalman gains consistent with an adaptive weighting strategy (see figure 10).

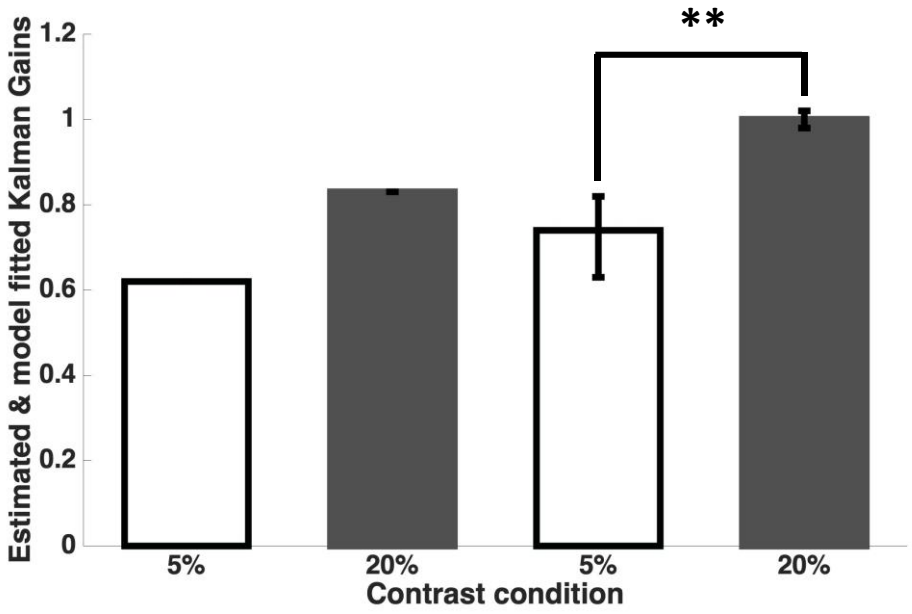


Figure 10. Estimated and model fitted Kalman gains for experiment two. Here, we observe the clear and indeed significant differences in model fitted Kalman gains between contrast conditions in experiment two. Estimated Kalman gains for the 5% were 0.62 while in the 20 % we calculated gains at 0.83. This was a slightly different outcome to our model fitted Kalman gains. In the 5% contrast we record a Kalman gain of 0.74 while in the 20% condition we recorded a Kalman gain of 1.02. This result provides an interesting comparison with our predicted Kalman gains for experiment two and if correct predicts serial dependence in only the 5% contrast condition but not in the 20% condition and importantly an apparent effect of the ratio between proximal and distal variance and an adaptive weighting process

Mean model fitted weights

Analysis of model fitted weights were consistent with our differences in Kalman gains between contrast conditions and indicate a more distributed weighting in the 5% contrast condition than in the 20% contrast condition (see figure 11 and table 3). Mean weighting figures and upper and lower bound 95% confidence intervals for the 5% percent contrast condition indicated a distributed weighting albeit with some inter individual variations and are illustrated below in figure 11and reported in detail in table 3.

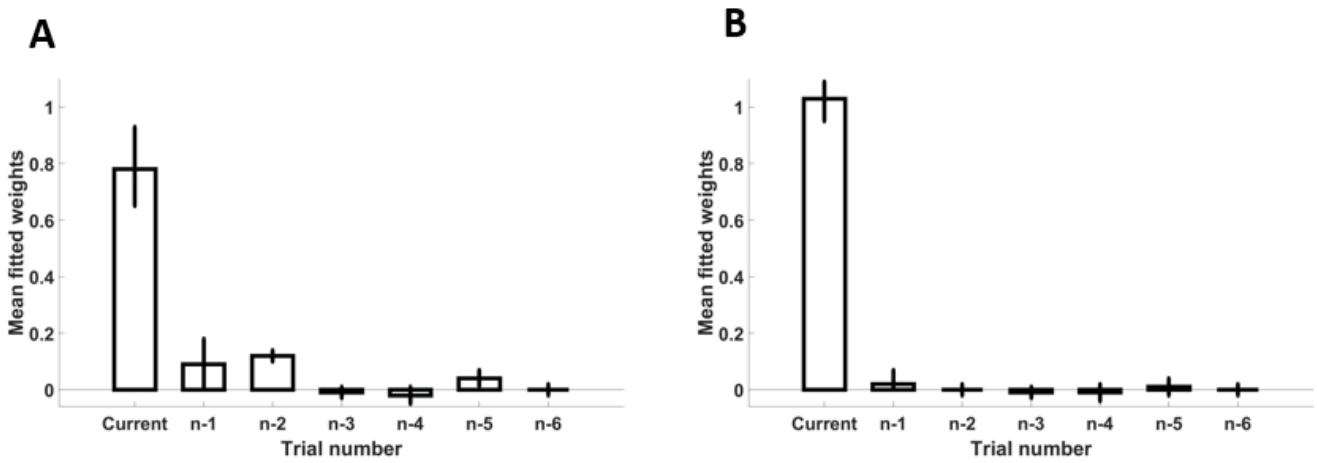


Figure 11. Mean model fitted weights for experiment two. Sub figures A & B illustrate mean model fitted weights in the 5% (A) & 20% (B) contrast conditions on the current and previous 6 trial orientations. Here again we see evidence of an adaptive weighting of information over time modulated by the ratio of proximal to distal variance. In the 5% contrast condition, weights are distributed over the current and previous two trials while in the 20 % contrast weighting is confined entirely to the current trial orientation .While it is correct to note that our model fitted analysis is only fitting a model to data it does appear that our both our models provide support for an adaptive weighting strategy and essentially concur with Kalman gains on the way weighting is distributed over conditions.

Table 3 Current/N back trial fitted weights and 95% confidence intervals for experiment two

	5% Contrast		20% Contrast	
	Weight on trial	CI (LL, UL)	Weight on trial	CI(LL,UL)
Trial				
Current	0.78	[0.64, 0.93]	1.03	[0.95, 1.10]
N-1	0.09	[0.02, 0.19]	0.00	[0.00. 0.10]
N-2	0.12	[0.06. 0.18]	0.00	[-0.05, 0.05],
N-3	0.01	[0.03, -0.03]	0.00	[-0.03, -0.03]
N-4	-0.02	[-0.04, 0.04]	-0.01	[-0.05,0.05]
N-5	0.04	[-0.06, 0.15]	-0.01	[-0.05, 0.03]
N-6	0.00	[-0.04, 0.07]	-0.02	[-0.06, 0.03]

Serial dependence predictions for experiment two

Our Kalman gains and model fitted weights allow us to make specific predictions about the level of serial dependence we might observe in experiment two-albeit with some slight ambiguity between estimated and model fitted Kalman gains. Estimated Kalman gains of 0.62 (5% contrast) and 0.83(20% contrast) mean that we should observe high serial dependence for both conditions. Translating these figures to our serial dependence regression slopes we can expect a slope of 0.38 (1-K(0.62)) for the 5% contrast condition. This would be a huge amount of serial dependence. Model fitted Kalman gains make a different prediction about serial dependence. In the 5% we should observe a regression slope of 0.26 (1-K(0.74)) which is still a large amount of serial dependence but in the 20% condition we would expect no serial dependence (1-K(1.02) and in fact we might even expect some repulsion effects as opposed to bias towards previous orientations. Fitted weights also provide a prediction about serial dependence on our n back analyses. With a weight of 0.78 on the current trial for the 5% contrast condition and some weight on the previous 2 n back trials we would expect some serial dependence going back over trials. In the 20% contrast condition, the weight was around one on the current trial. In this case, we would expect no serial dependence on the immediately previous trial orientation and no serial dependence to any previous n back previous trial orientation. In summary, predictions made by both Kalman gains and fitted weights state that we should observe different amounts of serial dependence between contrast conditions thus supporting the adaptive predictive coding account of the phenomena.

Individual serial dependence illustrations.

Our individual serial dependence error plots for four participants from experiment two (see figure 12) provide an early indication that the adaptive account of serial dependence may be supported. Here, while we do not carry out statistical tests we see the clear difference in serial dependence between conditions, with on some occasion a very high level of serial dependence recorded in the 5% contrast condition. Another interesting aspect of our individual plots is the visible relationship between error variance (proximal variance-the width of the shaded regions) and serial dependency which is especially evident in the 5% contrast condition. When we have a low level of proximal variance participants generally respond at the current trial orientation. However, when we have a high level of proximal variance participant's error generally becomes serially dependent on the previous trial orientation. In some participants (A) the relationship is very strong with a regression slope coefficient of 0.26 while in others the effect is less strong. For example in sub figure B the regression slope coefficient is at 0.11. Also note in sub figure (A) the subjects highly variable response errors which can be considered to be the effects of stimulus visibility in both the 5 & 20% contrast conditions i.e. proximal variance This is especially interesting as this is the only subject in which we observe serial dependence in both contrast conditions. A further interesting aspect of our individual data plots is that we again observe 'negative' serial dependence in sub figure C. As we observe it again it is increasingly unlikely that this is not down to statistical errors or chance but that perhaps a different perceptual mechanism or strategy is at work.

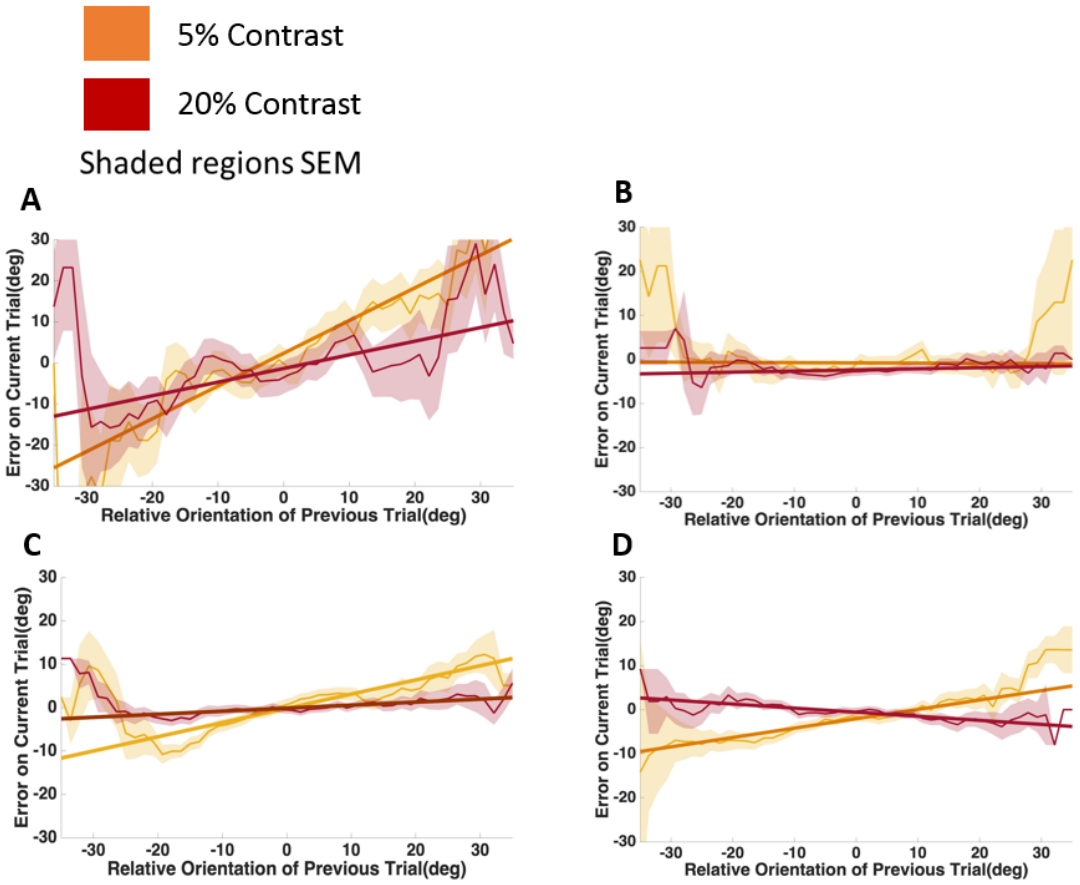


Figure 12. Four individual serial dependence plots. Here we can observe a possible effect of proximal variance on serial dependence. In participants with larger SEM (sub figures A & C) there appears to be larger amount of serial dependence-note in sub figure A we see this effect strongly. Also note the quite strong negative serial dependence in subfigure D the reason for this is unclear especially in light of the large individual differences in responses. Indeed, this is a recurring issue in our data with different participants responding in quite different ways to the trial orientations.

Serial dependency main analysis.

We hypothesised that the magnitude of serial dependence is adaptively modulated by the ratio of proximal and distal variance in our stimulus and commensurate to the level of Kalman gains. Here we found this to be the case and report significant differences in regression slope coefficients between contrast conditions with 5% (M=0.20 SD=0.20, CI [0.63, 0.88]) and 20% (M=-0.02, SD=0.99, CI [0.99, 1.05]) contrast conditions t (9) =4.791. *p*=.001, strongly indicating an adaptive integration strategy (see figure 13 below). While, the amount of serial dependence we recorded in the 5% contrast condition is slightly below the level predicted (see figure 13 for an illustration of predicted versus actual values), the fact that the magnitude of serial dependence changed under different levels of proximal and distal variance and were generally commensurate to our levels of Kalman gain supports our adaptive hypothesis. The effects of proximal variance can be seen in figure 13 below. In this plot, the green shaded standard error region (20 % contrast) is narrower than the purple shaded region (5%) as this is the only aspect of the study that was manipulated it indicates that stimulus variability plays a prominent role in the integration of past information. Results also show just how difficult it is to induce the visual system to integrate past information into current perceptions. To produce a reasonably high level of serial dependence in experiment two, we had to dramatically reduce distal variance from experiment one. This results again contradicts Fischer & Whitney (2014) although it must be said we did not directly replicate their experiment. In the next section of our results we do though report a more direct replication that was also carried out.

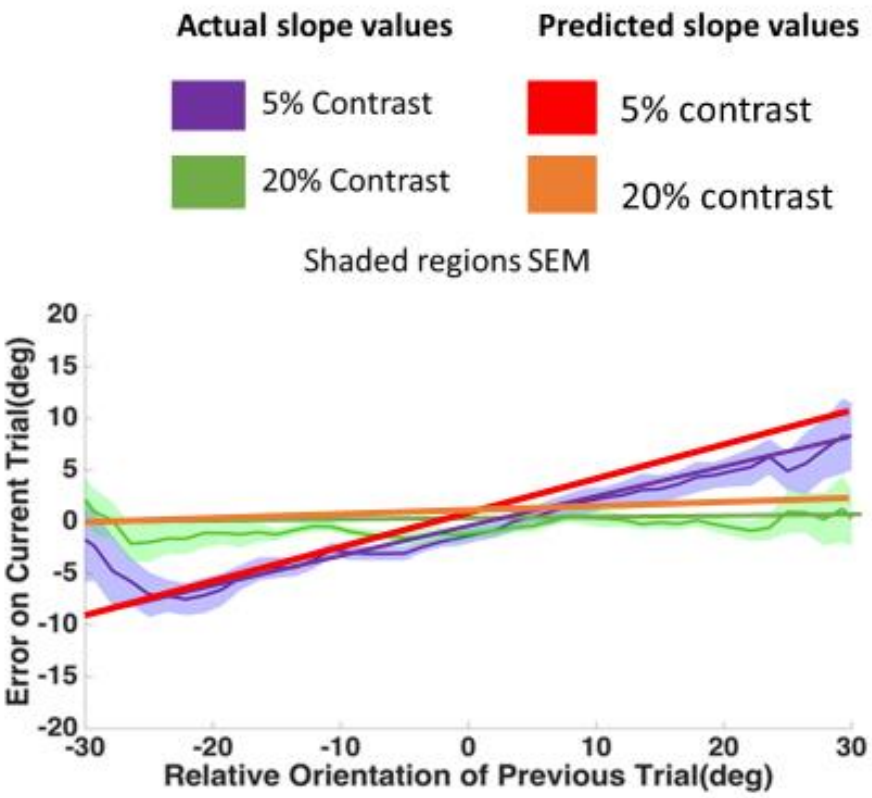


Figure 13. Group error versus relative orientation plots. Here we observe the significant difference in regression slope coefficients between conditions. The slope coefficient in the 5% reported a slope of 0.20, indicating a high level of serial dependency while in the high contrast (20%), low proximal condition, the slope is zero indicating no serial dependency. This is a vital result in terms of our analysis of an adaptive versus a fixed hypotheses. A significant difference in slope values indicates that when the amount of distal variance is reduced and measurement uncertainty increases, carrying higher proximal variance, as recorded in the 5% condition in this experiment, stimulus history influences participant's estimates of the current stimulus value and we observe a large amount of serial dependence.

Serial dependence n back analysis

As with our main analysis of serial dependence in our n back analysis our modelling provided a relatively accurate prediction of serial dependence. Results were especially consistent with our model fitted weights. As predicted, findings indicated that we observed serial dependence towards the previous six trial orientations in the 5% contrast condition and no serial dependence recorded in the 20% contrast condition with regression slope coefficients of , n-1 0.19, n-2, 0.12, n-3 0.09, n-4, 0.09, n-5, 0.07 & n-6 0.03 and 20% contrast n-1, 0.00 , n-2, 0.00, n-3, 0.00, n-4, -0.01, n-5, -0.01 & n-6, -0.01 (See figure 14 and table 4). Note here that serial dependence appears to persist longer than model fitted weights suggested in the 5% contrast condition and that at least some subjects contrast appear to exhibit negative serial dependence from around the n-3 trial in the 20% contrast condition (see fig 14 below). The repeated occurrence of a small amount of negative serial dependence as observed in experiment one again represents a somewhat anomalous as aspect of the data. As it is repeated from experiment 1, it suggests that it is not down to chance or statistical error. However, crucially, the differences in the magnitude of serial dependence between conditions support our adaptive hypothesis of serial dependence and are consistent with ideas of predictive coding.

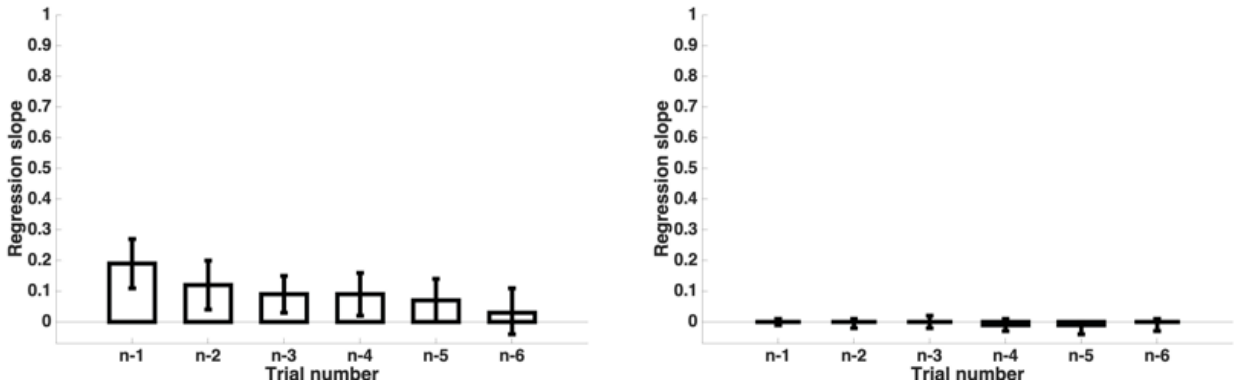


Figure 14. Serial dependence magnitude towards trial presented 1-6 trial back. In the 5% contrast condition regression slope coefficients report a slope of 0.19 for the n-2 trial decreasing over n-backs to a regression slope coefficient of 0.07 in the n-6 trial. A linear decrease in regression slope coefficients is consistent with results from our fitted gains and weights which indicate an influence of trial history over a number of previous trials decreasing over time in this condition. This persistence of serial dependence over time is not observed in the 20% contrast condition, in which as expected, no serial dependency is observed and in data fact hints at a slight repulsion effect towards previous trial orientations.

Table 4 Serial dependence N back trial analysis for experiment one

	5% Contrast		20% Contrast	
	Regression slope	CI (LL, UL)	Regression slope	CI (LL,UL)
Trial				
N-1	0.19	[0.11, 0.27]	0.00	[-0.01, 0.01]
N-2	0.12	[0.04, 0.20]	0.00	[0.01, 0.01]
N-3	0.09	[0.02, 0.15]	0.00	[0.01, 0.02]
N-4	0.09	[0.00, 0.14]	0.00	[-0.01. 0.03]
N-5	0.07	[0.04. 0.13]	-0.01	[-0.02, 0.02]
N-6	0.03	[-0.04, 0.10]	-0.04	[-0.05, 0.02]

3.6. Supplemental experiment.

Previous serial dependence literature has always reported at least some amount of serial dependency using an orientation judgment paradigm (Fischer & Whitney, 2014; Liberman et al., 2016; St. John-Saaltink et al., 2016). However, we found none in all but one experimental condition in one experiment. Although this finding was well predicted by our modelling, we felt it warranted further investigation. This was especially true in regard to our replication of Fischer & Whitney (2014). This is because, while it is true we conceptually replicated Fischer & Whitney (2014) we did not attempt a direct replication of their design. Due to the strength of Fischer & Whitney's (2014) methods and results it was important that we verify our results by performing a more direct replication as some aspects of the way Fischer & Whitney's (2014) stimuli were presented differed from our stimuli presentation. The differences here, applied mainly to the reduction of spatial frequency of the stimulus in their experiment compared to ours, a larger Gabor stimulus, a stronger inter trial noise mask in the same position of the stimulus and the screen positions of the Gabor (centre and periphery). The general effects of the differences between Fischer & Whitney (2014) can be considered to have the effects of a) making the stimulus more blurred and b) in trials presented in the periphery moving the stimulus outside of the foveal representation and thus increasing uncertainty. Given both factors could have increased what could be considered proximal variance under the principles of the Kalman filter it may have explained our non replication i.e. our stimulus had lower proximal variance values meaning participants were more likely to respond to the current trial orientation in our experiment. We now provide methods and results.

3.6.1 Methods.

Participants.

The supplemental experiment had 5 participants (3 female, mean age 26, range 19-41). Participants had not taken part any of our previous experiments and were recruited from St Andrews SONA participants recruitment database.

Supplemental experiment design and task.

Gabor patches were presented at 0.25 contrast has a radius of 10 visual degrees, a spatial frequency 0.33 cycles per visual degree and a 1.5 s.d Gaussian contrast envelope. Between presentations of Gabors a noise patch was presented at the same location as the Gabor. Noise patches comprised white noise smoothed with a 0.91 Gaussian kernel and windowed in a 1.5 s.d Gaussian contrast envelope. The experiment had two conditions. In condition one (peripheral), Gabors were presented 6.5 visual degrees to the right of fixation which was a 0.5 diameter dot (visual degrees) was presented in the centre of the screen. In condition two (central), Gabors were positioned centrally. In both conditions, Gabors were presented randomly between 0-360°. The orientation adjustment bar used to signal the participant's perceived orientation was a 0.61° wide white bar at 4° in length. The adjustment line was always positioned at the same location as the Gabor trial location. The response bar changed its starting orientation randomly on each trial. Each participant completed four blocks. Each block had 52 trials in each condition (peripheral and central). This meant each participant completed 208 trials with the Gabor positioned in the periphery and 208 with the Gabor positioned in the centre and 416 trials in total.

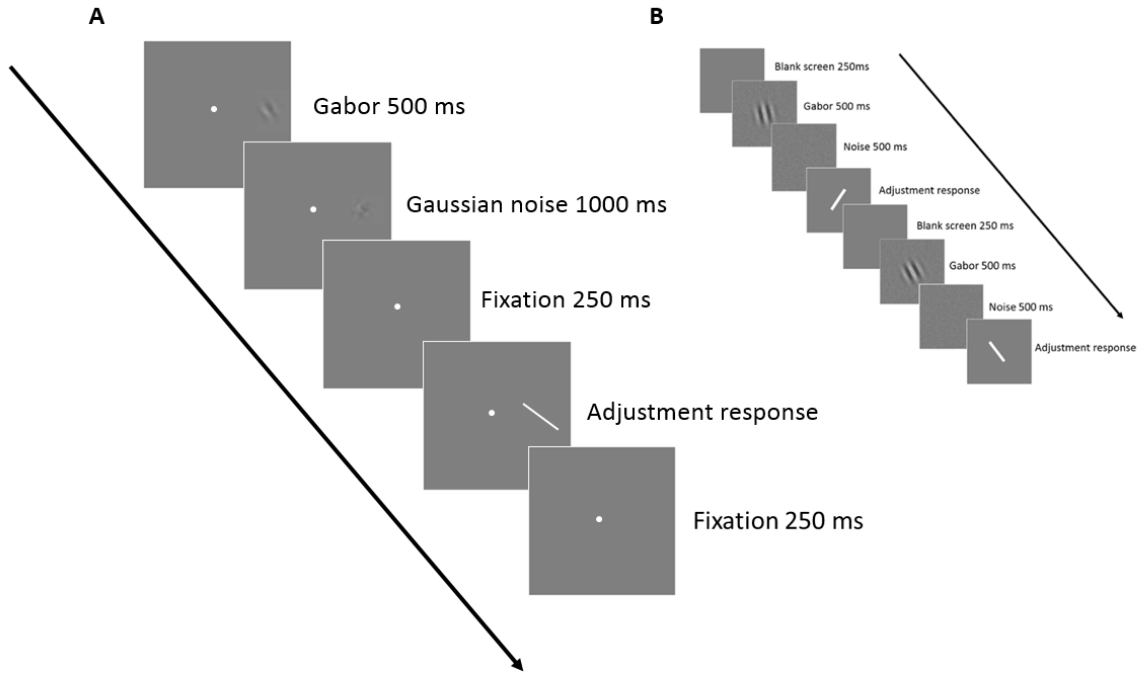


Figure 15. Design and task for the supplemental serial dependency experiment and comparison with our design and task from our previous serial dependency design and task. Sub figure A illustrates the design and task for our supplemental serial dependency experiment which is closer to Fischer & Whitney's (2014) design. Here, we can compare the important differences between our previously used more conceptual replication which it must be said combined to make the supplemental experiment qualitatively distinct. The reduction in spatial frequency had an effect of making the orientation of the Gabors more blurred and the increase in thickness of the adjustment line made for perhaps a more noisy response.

3.6.2. Analyses.

All analyses methods are identical to the previous main serial dependency analyses except we only include model fitted Kalman gains as predictors for serial dependence and do not include an n back analysis.

3.6.3 Results.

Model fitted Kalman gains.

No effect of screen position on Kalman gain was observed in our analysis of model fitted Kalman gains with periphery, 5% (M= 1.01 SD=0.07) and 20% (M=0.95 SD =0.06) t (4) =-1.04. p=0.232 with upper and lower bound 95% confidence intervals for the peripheral condition, CI [1.11, 0.91] and central condition CI [1.03, 0.88] (see figure 16). Importantly, figures for Kalman gain are at or around one indicating no impact of trial history.

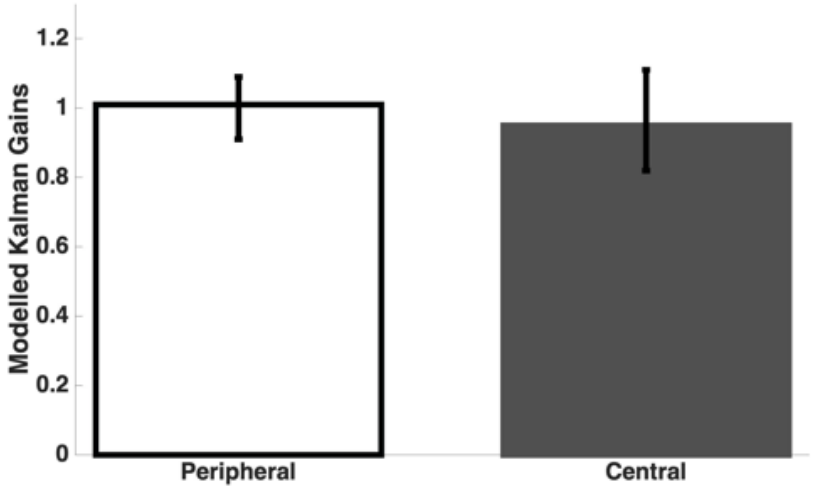


Figure 16. Modelled Kalman gains supplemental experiment. Here we can observe the similarities in model fitted Kalman gains in our two experimental conditions. Kalman gains of 1.01 for the peripheral condition and 0.95 for the central condition. This result bears marked similarities with our main serial dependency in experiment one which also presented orientations at random orientations between 0&360 degrees, indicating at such a high level of distal variance little if any weighting to previous stimulus values in participant's responses.

Serial dependence modelling predictions for our supplemental experiment.

Our model fitted Kalman gains were again close to one for both peripheral and central conditions. Based on this level of Kalman gain we would expect to see little if any serial dependency at the group level. However, given our lower bound confidence intervals indicating a slightly lower Kalman gain in some participants it is possible that some individual participants may exhibit a small amount of serial dependence.

Individual serial dependence plots.

Before providing group level results, in keeping with our previous results sections, we again provide an illustration of individual serial dependence plots. On this occasion, we again observe a similar result to our main experiment one, which also presented randomly oriented Gabors. Again, all regression slope coefficients are very close to zero and indeed there seems to be even more of the negative serial dependence or repulsion effect we observe in our previous experiments. Such results again lend weight to the effects of very high distal variance on the way participants respond to changes in stimuli over time. Namely, that when presented with a highly variable stimulus participants ignore past values and base responses on the current trial orientation.

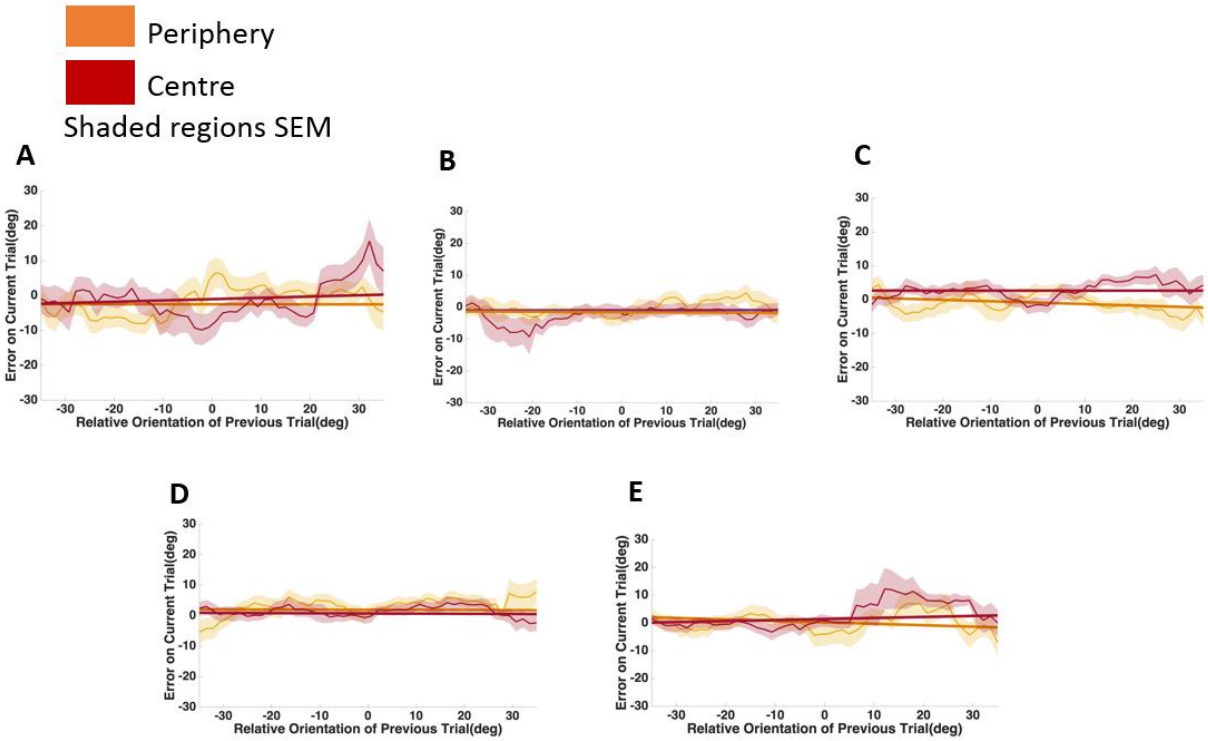


Figure 17. Serial dependence plots for all five individual participants. Here we again observe a similar result to our main experiment one which also presented randomly oriented Gabors. All responses are centred on the current trial orientation indicating no effects of trial history in responses.

Group serial dependence analysis.

Results from our analysis of serial dependency in our supplemental experiment are commensurate with our Kalman filter model predictions from this experiment and consistent with our initial findings in relation to serial dependency. Model fitted Kalman gains at or close to one again predicted very little or no serial dependency and no significant differences in regression slope coefficients between our peripheral and central conditions. This is exactly what was observed with peripheral (M= -0.020, SE=0.01) & central (M=0.01, SE=0.01), t (4) = -1.859, *p*=0.137. Upper and lower bound 95% confidence intervals were peripheral condition CI [0.01, -0.08] and central condition CI [0.04, -0.02]. It is worth noting two aspects of our data in this analysis. In this experiment, if anything we get even less serial dependence than in our previous experiments with a randomly presented stimuli and again we appear to observe a very slight 'negative' serial dependence or a repulsion effect in both peripheral and central conditions which can be observed in figure 18 below**.**

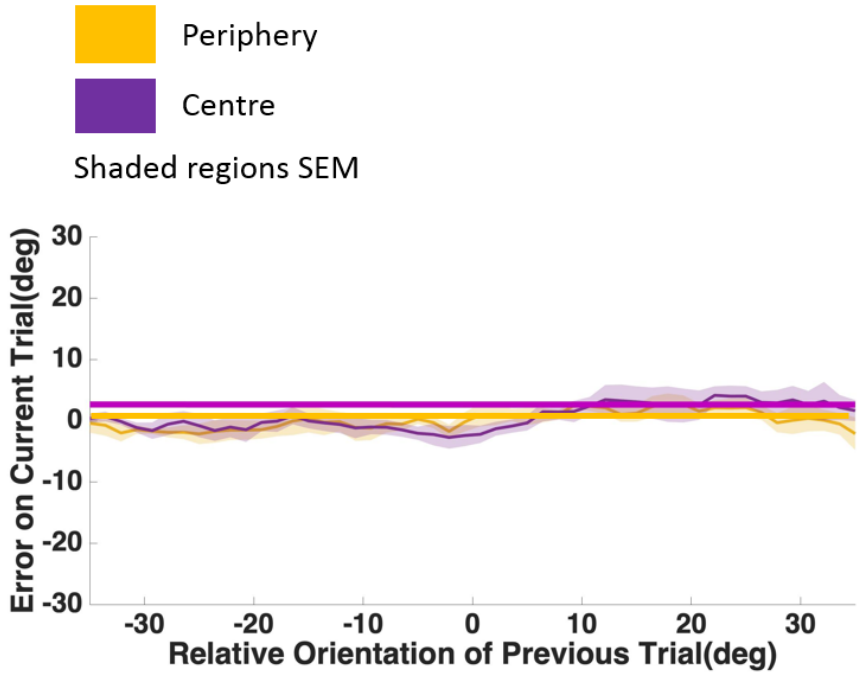


Figure 18. Supplemental serial dependency error versus relative orientation plots. We again fail to replicate Fischer & Whitney (2014) with results consistent with our initial findings that report no serial dependency. Both the peripheral and central condition report regression slope coefficients actually very slightly below zero.

3.7. Discussion and conclusion.

The current experimental chapter set out with very well defined goals with clear experimental hypotheses. We hypothesised that the magnitude of serial dependence is adaptively modulated by the ratio of proximal and distal variance in our stimulus and commensurate to the level of Kalman gain. Based on the overall findings, we found this hypothesis to be correct and our results over all are entirely consistent with ideas of predictive coding's ideas of precision weighting of prediction errors and well modelled by the Kalman filter. Furthermore, aside from what can be considered a novel overall finding within a well-defined theory our methods also provide some interesting techniques that could be used in future studies that examine serial effects and predictive coding. However, the study does raise some interesting issues that require further explanation. Not least, the failure to replicate Fischer & Whitney (2014). We now discuss our study beginning with the validity of our methods.

General methods

Perhaps the most important and indeed interesting aspect of our methods was the way we provided a measure of proximal variance and the implications it produced. To reliably calculate Kalman gain a measure of this quantity is required. While distal variance was pre-set in both experiments and a known quantity, proximal variance is more difficult to measure as it is very much participant specific and difficult to quantify. Indeed, perhaps for this reason other papers that have linked the Kalman filter with how estimation errors are weighted have commonly ignored the part of the equation for Kalman gain that requires a figure for measurement variance (Burr & Cicchini, 2014; Cicchini et al., 2016). However, although potentially problematic, significant and consistent results from our proximal variance calibration experiment showed that our manipulation of contrast and our analysis methods in this regard were successful. Crucially, in all participants, the variance of fitted cumulative psychometric functions was higher in the 5% contrast condition than the 20% contrast condition indicating a more variable internal measurement of the 5% contrast condition than the 20% contrast condition.

There are a number interesting features of the data provided by our proximal variance experiment. One is simply just how good vision is. In the 20% contrast condition all participants recorded a low level of proximal variance. This indicated that they could all see the stimulus well and respond accurately to the orientation. Given that we used noise masks that are known to make the perception of a stimulus more variable (Breitmeyer & Ogmen, 2000) and overall the appearance of the stimuli was qualitatively not totally clear this was somewhat surprising. Even in the 5% contrast condition, in which it must be said was very hard to see the stimulus, some participants were still actually quite good at judging the orientations. Our proximal variance figures also quantified how vital raising the level of proximal variance over the level of distal variance is for allowing predictions to be more useful than the current input. Our figures showed that in experiment one, distal variance was much

larger than proximal variance and here Kalman gains were all close to one indicating that we would not expect to observe any input of past information and this is exactly what we observed. In experiment two, we reduced the level of distal variance drastically (from 2700 deg^2 to 133 deg^2). This meant that for most participants their proximal variance was higher than the distal variance. Only when this occurred, did we observe any serial dependence. Proximal variance figures also show just how hard it is to reduce the level of distal variance below that of proximal variance. In experiment two, we had to make our stimulus sequence highly correlated and our stimulus bordering on the invisible to reduce proximal variance below distal variance. Finally, our proximal variance figures can be considered to be reasonably accurate if a little high. We can ascertain this from experiment two. This is because the predicted Kalman gains of 0.62 (5% contrast) and 0.83 (20% contrast) predicted more serial dependence than we actually observed. In fact the model fitted Kalman gains which had higher Kalman gains 0.74 (5% contrast) & 1.02 (20% contrast) were a better predictor of serial dependence. Overall, it could be considered that our methods for introducing and estimating proximal variance, while not perfect are a useful means to quantify uncertainty and could be applied in many areas of visual research that study the effects of uncertainty including predictive coding and serial dependence.

An additional aspect of our methods that represents an advancement on many previous predictive coding studies is the insertion of predictability into our stimulus sequences. Previous predictive coding studies have largely used cues or associative type relationships to induce predictability (Kok, Jehee, & de Lange, 2012a; Kok & Turk-Browne, 2018; Summerfield & Egner, 2009; Summerfield & Koechlin, 2008). In these studies trials are statistically independent from each other and indeed follow the randomly presented type of trial sequence in we used in experiment one. However, in experiment two we changed this to make trials statistically dependent to some extent on one another by inserting a Gaussian random walk into the trial sequence. A Gaussian random walk is a one dimensional markov process in which the value at one time point is conditionally dependent on the previous time point. Although the aim of inserting the random walk was mainly to reduce distal variability, in predictive coding theory, the nature of the internal models is such that they represent conditional relationships between events and stimuli over time (Clark, 2013; Friston, 2010; Thornton, 2014). The fact that we observed differences in behaviour indicating use of past trial history when trials were correlated (serial dependence) supports these ideas and the increased use of statistically dependent trials in future experiments. Indeed, this is an area in which the current thesis expands upon in chapter 5.

Modelling analyses.

The use of trial sequential regularities is key to one of the two models we used in the current study, namely the Kalman filter which was able to predict well the magnitude of serial dependence we observed. Predicted Kalman gains, based on proximal variance calibration figures for experiment two were 0.62 for the 5% contrast and 0.83 for the 20% contrast. Model fitted Kalman gains were 0.74 for the 5% and 1 for the 20% condition. The difference in Kalman gains for the 20% between models fitted gains and predicted Kalman gains is an interesting aspect of our data which could mean that two different predictions about the magnitude of serial dependence in the 20% contrast condition were possible. According to the model fitted gains that produced a gain of 1 we would not expect to see any integration of previous stimulus values but according to on the gain level of 0.83 derived from the proximal variance calibration experiment we might expect to see an influence of previous stimulus values in participants judgments of stimulus orientations in our integration analyses. Results from our serial dependence analysis supported data from our model fitted Kalman gains. In the 5% contrast condition we observed regression slope coefficients of 0.21 while in the 20% contrast condition a coefficient of zero was produced. As we have mentioned, this may be due to an over estimation of proximal variance in our proximal variance experiment. In summary both predicted and model fitted Kalman gains provided reasonable predictors of serial dependence with model fitted Kalman gains more accurate. This result was also commensurate with results from our model fitted weights which predicted serial dependence to n back trials to a certain extent.

Model fitted weights predicted the level of serial dependence to n back trials in both experiment one and two. In experiment one, model fitted weights of at or close to one for both the 5% and 20% contrast predicted no serial dependence to any of the 6 included n back trials. This is exactly what was observed. We found no serial dependence to any analysed trial. For experiment two, though another outcome was predicted. In experiment two, model fitted weights indicated that we would observe serial dependence in the 5% contrast condition stretching back three trials and again that we would observe no serial dependence in the 20% contrast condition. Again, our modelling analysis was broadly correct. We did not observe any serial dependence in the 20% contrast condition but did record it in the 5% contrast condition and in fact recorded serial dependence to the previous six trials degrading as a function of time. This aspect of our results was consistent with previous serial dependence studies (Bliss et al., 2017; Fischer & Whitney, 2014; Liberman et al., 2016) to some extent but as we did not observe any serial dependence in the 20% contrast condition does flag some important questions in regard to our general findings in regard to serial dependence literature.

Serial dependence, predictive coding and a failure to replicate Fischer & Whitney (2014).

Findings from our analysis of serial dependence support a predictive coding account of temporal integration. Although we cannot be certain the adaptive weighting observed in the current study applies to prediction errors, as we do not have access to the underlying activity of the brain, results all point to a predictive coding account of precision weighting. When stimuli are more uncertain and contain at least some amount of predictability, participants responses exhibited integration of past stimulus history. This result is broadly consistent with more recent findings indicating that serial dependence is modulated by viewing conditions in the form of distance to the foveal repression although again we did not observe such an effect (Fritsche, Mostert, & de Lange, 2017) and probabilistic information in trial sequences (Bergen & Jehee, 2019) which can be considered proximal and distal variance related respectively. Importantly, this was not observed during experiment one which presented random orientations.

Our interpretation of why we find no serial dependence to randomly changing stimuli is that quite simply with a randomly changing stimulus there are no relationships in the stimulus sequence for the brain to model. In this case, and what happens in predictive coding and the Kalman filter is that the current prediction error should be weighted to its full extent as the past carries no useful information. In this case, we would expect responses to be at the current trial orientation and observe no serial dependence. While we do not actually measure how error is corrected per se in this experiment our data fits this explanation more than that provided by the fixed weighting account of simple perceptual averaging. Of course, the Kalman filter does apply a type of averaging but it is also reliant on the level of proximal and distal variance and the input of model based information and importantly is adaptive. However, despite the support from both our modelling and serial dependence analyses our results flag a serious issue in that we failed to replicate Fischer & Whitney (2014).

There are a number of potential reasons why we did not replicate Fischer & Whitney (2014). One reason may due to the individual nature of serial dependence we observed in relation to the small number of participants Fischer & Whitney (2014) tested. In our experiments, some participants were responses were almost always serially dependent while others seemed never to be serially dependent. While Fischer & Whitney (2014) tested 12 different participants in total, in all of their individual experiments they only tested 4 participants. If those four participants had been what we could consider high proximal variance observers, that we have shown tend to exhibit higher levels of serial dependence, this may explain at least some of the discrepancy between results. However, this would not explain their reporting of serial dependency for a random stimulus which we certainly did not observe. An explanation for this is that Fischer & Whitney (2014) stimuli produced a much higher level of proximal variance that could have produced serial dependence to a random stimulus but this is questioned as we did carry out a more direct replication that again failed to replicate. Finally, a

potential explanation is that participants used by Fischer & Whitey (2014) had a different underlying model of the stimulus and task demands. In some predictive coding models, there exists not only an model of the stimulus behaviour but also of the task structure (Friston, 2010). It is possible that the four participants tested by Fischer & Whiney (2014) when faced with a random stimulus did not followed a non optimal strategy and integrate across trials even though this would potentially lead to increased error.

Negative serial dependence or repulsion effects.

One final issue of our data to mention that lacks clear explanation is the existence of small amounts of negative serial dependence in some participants responses. Initially, we felt that this was an unreliable or random effect related to chance with a limited number of trials and the way orientations were presented in experiment one. To clarify, in experiment one, when presented with a random series of orientations, it may have been possible that participants simply responded to the current orientation with some error that could randomly have been towards the previous trial or away from the previous trial. If there were simply more responses away from the current trial orientation then this would appear as a repulsion effect but was in reality due to the random nature of orientation and associated responses with a limited number of trials. However, as we again observed such an effect in experiment two when orientations were heavily correlated and on n back trials this perhaps this explanation is perhaps not sufficient. An alternate explanation forward in the literature is that repulsion effects observed in serial dependence may be related to adaptation after effects.

While adaptation effects are usually observed after longer stimulus exposure (Kohn, 2007) than in the current experiment some previous studies have also reported it effects following sub second exposure lasting for several seconds (Fritsche et al., 2017; Kanai, et al., 2007). These studies attribute the finding to an interaction between perceptual and higher level post perceptual memory processes in which a memory trace causes a negative aftereffect on perception. However, in such studies the negative aftereffect was only observed in situations when trials had broad confidence intervals indicating low confidence in the perception of the stimuli. However, we find the opposite and observe more negative effects in the more visible 20% contrast condition to which participants were much more accurate in their response. Furthermore, this account would not explain why we seem to observe such individual variability in the level of repulsion effects. Overall, we must say, at present we cannot explain the reason for the negative serial dependence we observe and is worthy of further investigation.

Conclusion.

There are what can be considered three important results from chapter 3. First, temporal integration appears tuned to the variability of sensory measurements (proximal variance) and the behaviour of the stimuli of interest (distal variance). When proximal variance was higher relative to distal variance participants perceptual estimates were closer to previous stimulus values and when distal variance was higher relative to proximal variance estimates were closer to current stimulus values. Secondly, serial dependence is adaptive. In contrast to previous characterizations of the phenomena (Fischer, & Whitney, 2014), it is not pervasive functioning at a fixed level and instead adapts to level of variability in the behaviour of stimuli (distal variance) and the reliability of the observation of the stimulus in question (proximal variance) more consistent with ideas from Bayesian visou-motor literature (Denève, Duhamel, & Pouget, 2007; Knill & Pouget, 2007; Wolpert, Ghahramani, & Jordan, 1995;Wolpert & Flanagan, 2001). Lastly, while predictive information does play a role in our results we found temporal integration strongly favours the current stimulus input over previous information. Despite designing a stimulus that was extremely hard to see, it was very difficult to observe the effects of past stimulus history in participants estimates until we greatly reduced distal variance. Overall, results favour the predictive coding account of temporal integration with some caveats over the strength of the role of predictive information on perception except under very uncertain conditions.

Chapter 4. Adaptive correction of response error-The Kalman filter and step response functions.

4.1 Abstract.

Predictive coding rests upon the idea that the brain is constantly predicting the nature and origin of its incoming signals based on a system of internal probabilistic models. To maintain the behavioural validity of its internal models and the predictions generated, the brain relies on errors or prediction errors and between predictions and sensory information. Crucially, to understand how prediction errors mediate predictions it is important to consider how prediction errors should be weighted in relation to previous predictions and sensory inputs. If sensory information is underweighted the weight attached to prediction errors will be too low. This can lead to an over reliance on predictive information from the past making perceptions too slow to correct to change in the world. Alternatively, if sensory information is weighted too heavily the weight attached to prediction errors will be too high. In this case the weight attached to previous predictions will be too low making prediction errors overly sensitive to noise and other unimportant factors. Theoretically, obtaining the correct weighting corresponds to optimizing the Kalman gain in the Kalman filter. The general relationship between Kalman gain and error correction is that when Kalman gain is closer to one prediction errors are weighted more heavily, causing estimates to update faster. Alternatively, with a Kalman gain closer to zero prediction errors are weighted less heavily casing estimates to update more slowly. Previously, in chapter three, we provided general support for this idea by showing that serial dependence in perceptual estimates was adaptive to the ratio of proximal and distal variance and commensurate to the level of associated Kalman gain but this method concentrated on testing the adaptive versus fixed nature of perception as opposed to error correction per se. Here as a means to test further the idea of adaptive weighting of prediction error in a more direct way we manipulate proximal and distal variance in four conditions which use a step response type experimental design widely used in control theory to test the way estimating systems adapt error correction to sudden change (condition 1, 18.6 °step/5% contrast, condition 2, 18.6° step/ 20% contrast, condition 3, 35.7 ° step/5% contrast and condition four 35.7 ° step 20% contrast). We test the Kalman filter account of error correction against the fixed weighted average account at an individual and group level. At an individual level we examined the general relationship between proximal variance recorded in a separate proximal variance calibration experiment and Kalman gain. While results were not significant we did find a moderate negative correlation between proximal variance and Kalman gain that invites further investigation. At the group level we did observe highly significant results between proximal and distal variance conditions on Kalman gain thus supporting the predictive coding account of prediction error weighting and correction.

4.2 Introduction.

Classical theories of perception paint vision as a mainly passive process in which we receive input from the world and process information contained within these inputs as information travels from lower to higher visual regions and respond accordingly (Searle, 2015; Warren, 2012). While this view, was and is, still very prominent it does not adequately account for findings indicating that perceptions is also perhaps as much influenced by our previous experience and internal states as sensory information directly from the world. Historical ideas, such as analysis by synthesis (for a review see Yuille & Kersten, 2006) and perception as hypothesis testing (Helmholtz, 1863; Gregory, 1970) offered broad explanations of how past information influences perception but they were somewhat separated in both theoretical viewpoints and time. More recently, these ideas have been combined and formalised in modern day neuroscience by models of predictive coding (Rao & Ballard, 1999; Friston, 2012). The main idea of predictive coding is that a series of internal models of the external environment work to generate predictions about the most likely cause of incoming sensory inputs and neural activity patterns. Any errors between predictions and sensory inputs produce prediction errors which signal to the brain that something in the world might have changed and it should update its internal model. However, the key words to highlight in that sentence though are 'might have changed'. Because sensory information contains varying levels of noise and external stimuli exhibiting fluctuating levels of change applying a blanket weight to each prediction error would not be an optimal behavioural strategy and a way to determine the reliability of prediction errors is required.

How to weight prediction errors?

Determining the optimal way to weight for weighting prediction error is perhaps the most important aspect of predictive coding. According to predictive models, such as the free energy principle (Friston, 2002; 2018), if the weight attached to sensory information is too low then prediction errors might not be weighted sufficiently potentially leading to an over reliance on predictive information. This can lead to an over emphasis on past information and make perceptual judgments too slow to correct to change in the world. Alternatively, if we attach too high a weight to potentially unreliable sensory input the brain might attach too high a weight to our prediction errors. This could lead to an over reliance on potentially unreliable sensory information making perceptual judgments too quick to correct errors when a real change may not have occurred. In predictive coding literature, there are a number of individual models (O'Shaughnessy, 1988; Rao, 1999; Friston, 2010; Spratling, 2015) which provide differing explanations of how prediction errors are weighted for reliability in different circumstances in a number of visual modalities. However, to date there is no accepted tractable computational account of predictive coding that explains the way prediction errors are weighted in the visual temporal domain.

Adaptive versus fixed correction of estimation error-Ideas for experimental design from control theory- Step response function in conjunction with the Kalman filter and fixed weighed average models.

If the brain does use prediction error as means to update perceptions then in practical terms only accurately weighted prediction errors are relevant. Exactly how to ensure that weighting is accurate is by no means a problem confined to neuroscience. In Engineering and control theory, obtaining the correct weighting is identical to optimizing the Kalman gain in the Kalman filter as we have outlined in chapters 2 and 3. In the Kalman filter, when prediction variance is higher in relation to proximal variance a Kalman gain closer to one is produced. With a Kalman gain closer to one, prediction errors are weighted as more reliable and predictions update to a larger amount of the size of the prediction error. Alternately, when proximal variance is larger than prediction variance a lower Kalman gain closer to zero is produced. When Kalman gain is closer to zero prediction errors are considered less reliable and predictions update by a lesser amount of the prediction error. Of course, predictive coding's ideas about prediction errors are not the only way that the brain could update perception and correct error and the Kalman filter is not the only model by which this updating might occur.

In ideas of perceptual averaging and the related phenomena termed serial dependence (Fischer & Whitney, 2014), the concept of prediction error does not exist adaptive or otherwise. Serial dependence literature generally implies a simple stimulus averaging strategy we discussed in chapters 2 and 3. This strategy is identical to that performed in fixed weighted average models used in signal processing which provides our comparative integration model to the Kalman filter. In fixed weighted average models estimates updating occurs each time a new stimulus value is recorded and is independent of proximal and distal variance and the magnitude of Kalman gain. In summary, in fixed weighted average models all estimates contain at least some past stimulus history and updating occurs independently and regardless of stimulus or viewing conditions.

Previously, we examined the adaptive versus fixed weighting of prediction errors by behaviourally testing and modelling serial dependence with the Kalman filter (Kalman & Bucy, 1963) as our predictive coding model and the fixed weighted average model as our perceptual averaging model. This approach was successful in terms of providing support consistent with predictive coding's general adaptive weighting strategy. However, Fischer & Whitney's (2014) experimental paradigm while excellent for examining serial dependence, due to the way stimuli are usually presented and the way serial dependence is calculated means it is not the best medium to look at how error is reduced over time in a direct way. Serial dependence experiments commonly present stimuli in a way in which values either change randomly over time or are counter balanced between small numbers of values that change on every presentation (Corbett, Fischer, & Whitney, 2011; Fischer & Whitney, 2014). Because values change on every presentation it means that it is hard to see how estimation errors might reduce over time as they make multiple predictions and stimulus comparisons of repeated

values. Furthermore, serial dependence is considered as a bias towards previous values and calculated based on the error on the current response compared to previous stimulus values. This means the emphasis is always on the past rather than new stimulus values that are immediately relevant to behaviour. In predictive coding, the emphasis is not on the past but on reducing error in current and future predictions. With these issues in mind, perhaps a better experimental medium to observe how error is reduced to towards current values comes from ideas about step response functions.

The step response function-what is it, terminology and motivation

Step response functions are widely used in control theory for understanding the way estimating systems respond under different conditions. In technical terms, the step response function is the temporal response behaviour of the outputs of a general system such as a temporal filter or another estimator when the current inputs provided to it change suddenly and then stabilize (see Zumbahlen (2008) for an overview of the step response function and related terminology and figure 1 below for an illustration), (we actually used a step response function as an example of the response properties of both the fixed weighted average and Kalman filter models in chapter 2. In more basic terms, the step response function shows how a system reacts to a stimulus that remains stable for a time, jumps in value for an approximately equal time and then steps back to its original value. By presenting stimulus values that contain distinct stable and changing phases of equal length we aim to show how estimation error to current stimulus values reduces under different stimulus conditions in a more illustrative way than methods which present stimuli that change on every presentation and mainly look at the relationship to past values (see figure 2).

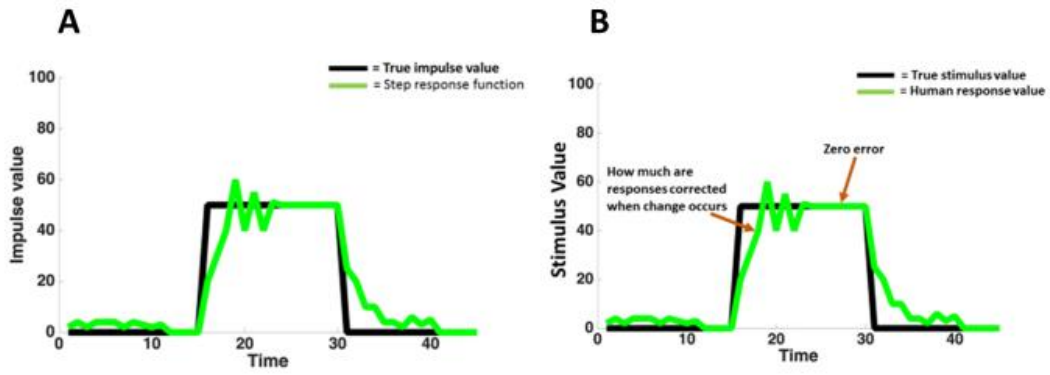


Figure 1 Step response function (A) & Extrapolation and application of the step response as a measure of the rate of error correction and time taken to reduce error to zero in humans (B). In sub figure A Here we see a basic step response function. Here the impulse values of the incoming signal in stimuli start at 0 for 15 seconds and then suddenly change to 50 and then remain at 50 for another 15 seconds before going back to 0 for the last fifteen seconds. Here we see the 'step response function' of a hypothetical filter model to the time behaviour of the impulses. The interesting points here are the time taken to reach maximum value often called 'rise time' and the time taken for the filter to reduce error to zero which is often called 'steady state' which provide useful terms in thinking about how humans might respond to change and reduce error in perceptual estimates. Sub figure B shows how we can translate the step response function to the way humans might correct error when responding to sudden steps in stimulus value. The impulse can be simply switched to being considered as a stimulus input, rise time can be considered the rate to which humans might respond to a sudden change in stimulus value and how they might weight the reliability of the stimulus on each input while steady state being considered as when the human estimate finally reached the true stimulus value.

Previous studies have utilised ideas about step response functions to good effect in a number of areas of the literature. One area of visual literature that has long used step response function type paradigms to understand aspects how perceptual systems respond to change is smooth pursuit eye movements (Carl & Gellman, 1987; Robinson, 1965). Smooth pursuit is one of the two ways the eye can voluntarily shift gaze (the other being saccades) which only occurs during the movement of the head and acts as a stabilising mechanism when viewing stationary objects (Thier & Ilg, 2005). The type of step response paradigms used in smooth pursuit commonly involve step changes in the velocity of a stimulus participants are asked to track (Carl & Gellman, 1987; Ono & Mustari, 2008; Robinson, 1965). Such step paradigms have seen wide utility and while a different type of task and stimulus then we intend to use in the current chapter such experiments illustrate the way sudden changes in stimulus value are a useful means to understand the way the visual system deals with change and error.

The use of a step change in velocity in smooth pursuit experiments has revealed a number of interesting findings about the phenomena which in some cases are similar to the aims of the current chapter. For example, step velocity paradigms have shown that smooth pursuit is modulated by an adaptive gain mechanism sensitive to measurement noise and the magnitude of the step in velocity (Brostek, Eggert, & Glasauer, 2017; Nuding et al., 2009; Tavassoli & Ringach, 2009). Studies using step velocity paradigms have also reported that after repeated exposure to steps in value smooth pursuit changes velocity and effectively starts to predict the step change and offering a way to ascertain how the brain learns to anticipate changes to reduce error in eye movements (Barnes & Asselman, 1991; Braun, Boman, & Hotson, 1996; Fukushima, Fukushima, Warabi, & Barnes, 2013; Zambrano, Falotico, Manfredi, & Laschi, 2010). Such findings, although present in a separate area of visual perception provide relevant findings to the current chapters experiment and importantly provide validity to the use of step response experimental designs in the study of visual perceptions. However, although interesting they are found in a completely different perceptual domain very early in visual processing and may not apply to more complex visual tasks involving a later perceptual estimate task. Perhaps a different area of the literature that makes use of step response experimental designs in a more similar way to our aims is visou-motor recalibration.

Burge Ernst & Banks (2008) examined the way reaching errors are corrected under the term 'visuomotor recalibration' using a step response type experimental design alongside ideas from the Kalman filter. Burge Ernst & Banks (2008) wanted to test how measurement noise and changes in the perceived position (mapping) of a stimuli affected the way participants adapted to changes in the value of the stimulus over a number of experiments. In experiment one, the aim was to investigate whether stimulus measurement reliability affected adaptation rate. Participants were asked to hold a stylus on a graphics tablet in which the hand and tablet were hidden from view. A target stimulus was presented at a random position for 500ms. The task of the participant was to try and match the point of the stylus to the position of the target stimulus on the tablet. Upon touching the stylus on the tablet the participants were given feedback (in the form of a Gaussian blob superimposed on a dashed circle) as to how accurate they were based on the difference between target and feedback location which was designed to provide a reference point as to how much participants should recalibrate the on the next trial. The reliability of feedback was manipulated by blurring the Gaussian blob in two conditions (σ=4°x4° & 24x24°). During the first sixty trials feedback stayed the same relative to the target but on the 61st trial the feedback 'stepped' 8.2° up and to the right (5.8° horizontally and 5.8° vertically) relative to the pre-step mapping, as indicated by the offset between the dashed circle and Gaussian blob while on the 121st trial feedback reverted to the original pre step position. The way participants corrected error following the step in feedback values in experiment one study revealed some interesting parallels with the Kalman filter.

Under the principles of the Kalman filter, when the stimulus measurement (proximal variance) is uncertain the adaption rate to a change in stimulus value should be slower than when it is more certain. This is exactly what was observed during the trials after the feedback stepped in position. In the 4° x 4° blur condition which provided a more concise appearing feedback measure, error reduced considerably quicker towards zero than the more blurred appearing 24x24° blur condition, in which error correction exhibited what can be considered a slower rise time and never achieved steady state during the 60 trials of the step phase. However, while interesting the manipulation of measurement variability alone can only partially test the ideas of the Kalman filter and to be sure that lagged error correction is not simply down to a more simple averaging type strategy than that proposed in the Kalman filter manipulation of distal variance is required. This was carried out in experiment two of the study.

Experiment two of the study, switched from looking only at what might be thought of proximal variance under Kalman filter terminology to examining both proximal and what could be termed distal variance (which they call mapping uncertainty). To do this, Burge Ernst & Banks (2008) asked participants to carry out an identical task to that in experiment one but introduced random walks of different standard deviations into the relationship between the reach endpoint of the stylus and the visual feedback position in addition to step in presentation of feedback and again manipulated the blur of the Gaussian blobs. This meant the experiment had four conditions that each had different combinations of what can be thought of as proximal and distal variance levels. Condition one had low distal variance and high proximal variance (σ walk = 0.9° × 0.9°; σ blur = 24° × 24°), condition two had high distal and high proximal variance (σ walk = 2.5° × 2.5°, σ blur = 24° × 24°), condition three had low distal and low proximal and low distal variance (σ walk = 0.9° × 0.9°, σ blur = 4° × 4°) while condition four had high distal and low proximal variance (σ walk = 2.5° × 2.5°, σ blur = 4° × 4°). This more complete manipulation made it possible to make more specific predictions about the potential nature of the way errors might be corrected than in experiment one.

If the Kalman filter approximated the response of participants then each of the four conditions should have in theory produced different levels of Kalman gain and therefore different rates of error correction. For example, condition four with high distal and low proximal variance would lead to the calculation of the highest level of Kalman gain and correct error fastest while condition one with low distal and high proximal variance should lead to the calculation of the lowest level of Kalman gain and correct error at the slowest rate. Condition two (high proximal and high distal) and condition three (low proximal and low distal) should calculate Kalman gains somewhere in the middle and correct error accordingly. Alternatively, if the fixed weighted average model account of error correction is correct then there should have been no difference in error correction strategy between conditions with rates being the same for all conditions. Results provided support for the Kalman filter like behaviour in participant's responses over the fixed weighted average account of error updating.

Results from Burge Ernst & Banks (2008) were almost entirely consistent with Kalman filter theory and provide good validity for the application of step response experimental designs in vision. The slowest to correct to adapt to step in feedback mapping was the high proximal and low distal variance condition one while the fastest was indeed condition four which has the lowest blur and the highest variance in the random walk. Conditions two and three were also as predicted by the Kalman filter in the middle in terms of rate of adaption. Such results run counter to the fixed weighted average account of error correction in that the levels of measurements and perceived stimulus variability played a strong role in the way subjects adapted to change. If participants were attaching an equal weight to all of the stimulus inputs then this effect would not have been observed. Such results and the excellent use of a step response inspired paradigm provide good support for the application of experimental designs in purely visual experiments and we combine and build upon these ideas in our current experiments.

4.2.1 Theoretical motivation and hypotheses

One potential explanatory model that tells us how we should weight prediction errors is the Kalman filter where establishing the optimal weighting is provided by the Kalman gain. In the Kalman filter, Kalman gain adapts to the ratio between proximal and prediction variance largely modulated by change in the world (distal variance). Evidence for Kalman filter like behaviour has been observed in the correction of visou-motor errors (Berniker & Kording, 2011; Denève, Duhamel, & Pouget, 2007; Wei & Körding, 2010; Zylberberg, Pouget, Latham, & Shea-Brown, 2017) and may translate well to explaining error correction in visual terms. However, in perceptual averaging literature the updating process is based on fixed weighting of stimulus values observed over time (Fischer & Whitney, 2014; Liberman et al., 2014). Here, as in chapter three we perform a proximal variance calibration and then test the Kalman filter and fixed weighted average model by incorporating experimental ideas from Burge Ernst & Banks (2008) and a stimulus design from Fischer & Whitney (2014) with those behind the use of step response functions in control theory. We aim to build in steps of varying size into the presented orientations of our Gabor stimuli over trials in a similar way to Burge Ernst & Banks (2008) and manipulate the measurement variability of our stimulus but here we manipulate contrast instead of blur. Based on these theoretical ideas and motivations we can provide hypotheses at a sub group level where we aim to assess the relationship between proximal variance and error correction in high and low proximal variance observes and at a group level where we examine the level of Kalman gain which acts as a proxy measure for error correction under different levels of proximal and distal variance.

At a sub group level, if the Kalman filter account of error correction is correct then participants with higher levels of proximal variance should show a larger negative relationship with Kalman gain than participants who have lower levels of proximal variance. This is because in the calculation for

Kalman gain (see chapter 2) as proximal variance increases Kalman gain goes down as sensory information is considered unreliable. At the group level, we can hypothesise that in conditions with lower distal variance relative to proximal variance we will record higher Kalman gains and observe faster error correction to changes in stimuli.

4.3 Methods

Ethics

All calibration and experimental procedures were approved by the University of St Andrews Teaching and Research Ethics Committee. All participants gave informed consent.

Stimuli design software and presentation details

In all experimental and calibration procedures visual stimuli were created in MATLAB (The Mathworks Inc, Natick, MA) and presented using PsychToolbox (Brainard, 1997).

4.3.1. Proximal variance calibration experiment.

In order to provide an estimate of proximal variance we first undertake a proximal variance calibration experiment. This experiment aims to quantify proximal variance in both our proximal variance conditions (5% and 20% contrast) for application in calculating Kalman gain

Participants

A total of eleven participants undertook the proximal variance calibration experiment. (7 females, mean age 22, range 19-43). Participants had not taken part in any previous experiments and were recruited as volunteers from the St Andrews SONA database.

Trial numbers and experimental blocks

Over trials Gabor orientations were presented in pairs in which the second Gabor was presented at 8 different orientations +/- (clockwise or anti clockwise) from the first Gabor in the pair. These were 1°, 3.33°, 5.66°, 8°, 10°, 12, 15° & 60°. Each block presented 5 trials at every orientation difference in both contrast conditions (5% and 20%) giving 40 trials per condition and 80 trials per block. Participants completed 3 blocks giving 240 trials in total. All Gabor patches had a radius of 8 visual degrees and had a spatial frequency 0.5 cycles per visual degrees and were positioned centrally. Noise following Gabors is also Gaussian white noise (SD=15.5 cd/m^2) and covered the whole screen with a spatial frequency of 0.15 cycles per visual degree. Stimuli were viewed at a distance at 57 cm. Fixations were also positioned centrally (see figure 3 for an illustration of our stimulus design and procedure).

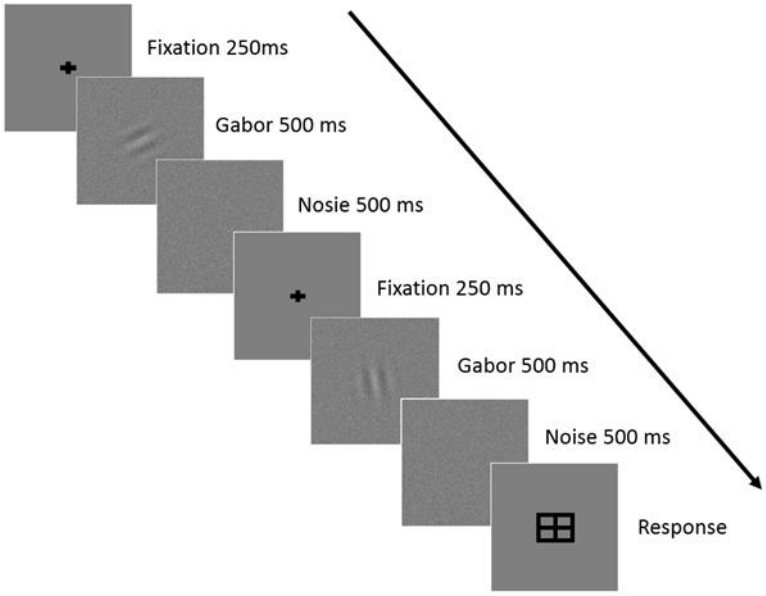


Figure 2. Proximal variance calibration experimental design and procedure. Each trial began with the presentation of a fixation cross in the centre of the screen for 250 ms. Then a randomly oriented Gabor patch was presented for 500ms at 5% or 20% contrast depending on the condition then a noise mask for 500ms. Next, a second Gabor was presented at one of 8 different angles ranging from 0-15° (1°, 3.33°, 5.66°, 8°, 10°, 12° , 15° & 60°) anti clock wise or clock wise from the first Gabor orientation and then a second noise mask for 500ms. The task of the participant was to fixate on the fixation cross and then discriminate whether the orientation of the second Gabor was clock wise or anti clock wise of the first Gabor orientation. This was signalled by pressing j for clock wise and f for anti-clock wise

4.3.2 Main experiment-testing the rate of error correction under different levels of proximal and distal variance over fours conditions.

Our main experiment aimed to manipulate proximal and distal variance in four conditions. We based this experiment on Burge, Ernst and Bank's (2008) experiments and observe the effects of each manipulation on the rate of error correction to steps in stimulus orientations of different sizes using Fischer & Whitney (2011) orientation judgment paradigm. Stimuli were Gabor patches. Gabor patches had a radius of 8 visual degrees and had a spatial frequency 0.5 cycles per visual degree. Noise following Gabors is also Gaussian white noise (SD=15.5 cd/m2) and covered the whole screen. The basic design and task is illustrated below (figure 4) before the experimental conditions for the manipulation of proximal and distal variance are provided

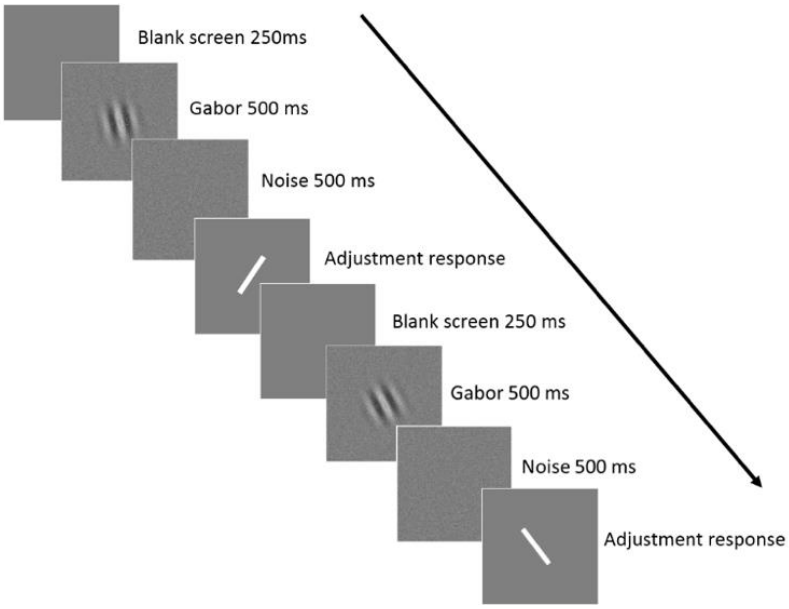


Figure 3. **Experimental stimuli in main experiments-design and procedure**. Each trial began with the presentation of a blank screen (inter trial interval) for 250ms. Next a Gabor patch was shown in the centre of the screen for 500 ms, then a noise patch was presented for 500ms then an adjustment response bar was presented. The task of the participant was to move the adjustment response bar to try and match the orientation of the Gabor they had just observed.

Participants

The same eleven participants who undertook our proximal variance calibration also undertook our main experiment.

Experimental conditions.

Condition one. Low distal variance and high proximal variance

In condition one, we aimed to produce low distal variance and high proximal variance. This was performed by presenting stimuli orientations in a step response function which was 18.6 degrees presented at 5% contrast. We chose 18.6° as it was 4 times the just noticeable difference (JND) of 4.65° observed in our proximal variance calibration experiment.

Condition two. Low distal variance and low proximal variance

In condition two, we aimed to produce low distal variance and low proximal variance. This was performed by presenting stimuli which stepped by 18.6 degrees and presented at 20% contrast

Condition three. High distal variance and high proximal variance.

In condition three, we aimed to produce high distal variance and high proximal variance. This was performed by presenting stimuli in a step response function which was 35.7 degrees presented at 5% contrast. 35.7 degrees was chosen as it was approximately 8 times the just noticeable difference (JND) of 4.65° observed in our 2AFC calibration procedure and 100% larger than the 18.6 degree step used in conditions one and two making for a logical comparison.

Condition four. High distal variance and low proximal variance

In condition four, we aimed to produce high distal variance and low proximal variance. This was performed by presenting stimuli in which the step size is 35.7 degrees and Gabors presented at 20% contrast.

Experimental blocking, trial order and numbers.

Conditions were presented in randomized blocks. Each block consisted of 45 trials which were broken down into three phases to comprise 3 steps and stabilization phases which were made up of 15 trials each (see figure 5 below). The first trial was a randomly presented oriented Gabor (0-360°). This initial trial orientation is repeated five times. On the sixth trial the orientation ‘steps’ in value by 18.6° or 35.7° depending on the condition and then the next four trials repeat this orientation then on the eleventh trial the orientation returns back to the originally presented orientation for a further 4 trials. At the end of this sequence of 15 trials another randomly chosen orientation and the whole step process begins again (see fig 5 for an example of the step response function). Participants completed six blocks of each condition over two one hour sessions which gave 1080 trials in total and crucially

allowed us to compare responses over 72 step and stabilization phases (1080/15) for each participant in each condition (see figure 5 below).

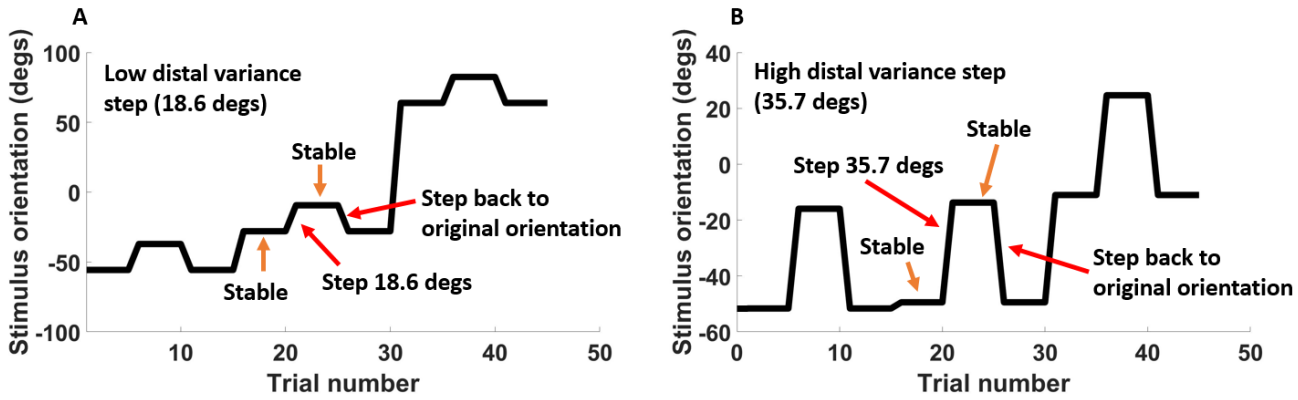


Figure 4. Stimulus step and stabilization phase illustration. Here we show how our stimulus behaves during the experiment for the two step sizes we use. Sub figure A shows the step function the low distal variance conditions (1&2) of 18.6 degrees. In this figure we see how the orientation starts at a randomly chosen orientation. This orientation repeats five times and then steps to a new orientation that is 18.6 degrees (plus or minus) of the previous orientation. The orientation then stabilizes at this orientation for five trials before returning to the original orientation for another five trial at which point a new orientation is selected and the process begins again. Sub figure B illustrates the high distal variance conditions (3&4) which follows the same process but in this case the step is larger at 35.7 degrees.

4.4 Analyses, statistical tests and equations.

Correction for potential ambiguity in responses using circular stimuli.

To correct for potentially erroneous recording of the orientations participants meant to choose we followed the same minimum angle difference correction to used previously n experiment three (see methods page 59)

Proximal variance calibration calculations.

To calculate proximal variance from our proximal calibration procedure we performed the same procedures and analysis as in chapter 3. To test for differences between contrast conditions we performed a Wilcoxon signed rank test as the data for proximal variance violated the assumption of normality required for parametric testing. We also report the coefficient of determination (r^2) as a measure of the proportion of variance shared by the two contrast conditions. This is calculated by dividing the Z value produced by the test by the square root of number of participants.

Correlations between proximal variance and Kalman gains at a sub group level.

In our analysis of Kalman gain and error correction between low proximal variance observers and high proximal variance observers we want to look at the correlation between proximal variance from

our proximal variance calibration experiment and the rate of error correction to the step in stimulus values. In this analysis, we utilise model fitted Kalman gains as a measure of the rate of error correction (1 being instant correction and 0 being never correcting) and assess the correlation between proximal variance and Kalman gain. As the range of variability in the 20% contrast condition was very narrow with all but one participant reporting variances between 10 & 17 deg^2 we felt it highly unlikely that we could observe any worthwhile correlational information from this data set. Therefore we only used the variance data from the 5% contrast condition which had a much larger range and offers a realistic opportunity to detect differences in correlations. As the 5% contrast variance data contained significant outliers the requirements the assumptions of parametric, Pearson r correlations are violated, therefore we use Spearman's Rho correlation as a measure of the monotonic relationship between proximal variance and Kalman gain. Here we also make a note in regard to the limitations of the statistical power of our correlations in this instant and the related *p* values from such tests. We recognise that due to the small sample size in each of the high and low proximal observer participant sets (n=4) steps any correlations indicating a relationship may not have sufficient power to be reliable. The same problem potentially applies to *p* values indicating whether any such correlations are significant. However, while recognising these limitations, we contend they still represent a useful if perhaps more illustrative measure of the relationship between measurement variability and error correction for this experiment.

Model fitted Kalman gains.

We computationally modelled participant responses at a range of Kalman gains using the same least squares model fitting procedure we had used in experiment 3 (see methods page 60). Again, it is important to reiterate that here Kalman gain is used as a measure of the rate of error correction to the step in stimulus values (1 being instant correction and 0 being never correcting). Our modelled analysis of Kalman gain produces a specific level of Kalman gain in each subject for each of our four conditions and therefore is suitable for statistical comparison. Statistical analysis of model fitted Kalman gains between conditions is carried out using paired sample t tests (Bonferroni-corrected). We report 95% confidence intervals of the mean.

Model fitted weight to the current and previous six trials.

As well as fitting Kalman gains to participant's data we also modelled participant responses with a fixed weighted average simulation model by 'best' fitting weights from the preceding 6 trial orientations to the current participant response.

Equations.

Equations for both the Kalman filter and a fixed weighted average model which we use in the current chapter are fully outlined in chapter 2, pages 42-46.

Simulated results.

Here we now provide a number of simulated step response function patterns we might observe in situations in which either a fixed weighted average model or Kalman filter most accurately describes our data.

Fixed weighted average step response simulations.

Below, in figure 6, we illustrate a number of step responses produced by a hypothetical fixed weighted average models with a weight of 0.5 to the current and previous (n-1) trial for each of our four experimental conditions. Sub figure A, shows a fixed weighted step response to condition one which has low distal and high proximal **(**18.6° step/5% contrast), sub figure B, shows how distal and low proximal variance condition two (18.6° step/20% contrast), the high distal and high proximal condition three (35.7° step/5% contrast) and the high distal and low proximal variance condition four. The key aspects in all of the above sub plots is that the rate in which the response corrects to the step and stabilization phases of the stimulus orientations is the same in all conditions. The estimate is always lagged in time, which while responsive to a certain extent, cannot adapt to suit clearer and more unstable conditions in which the stimuli changes by a larger amount.

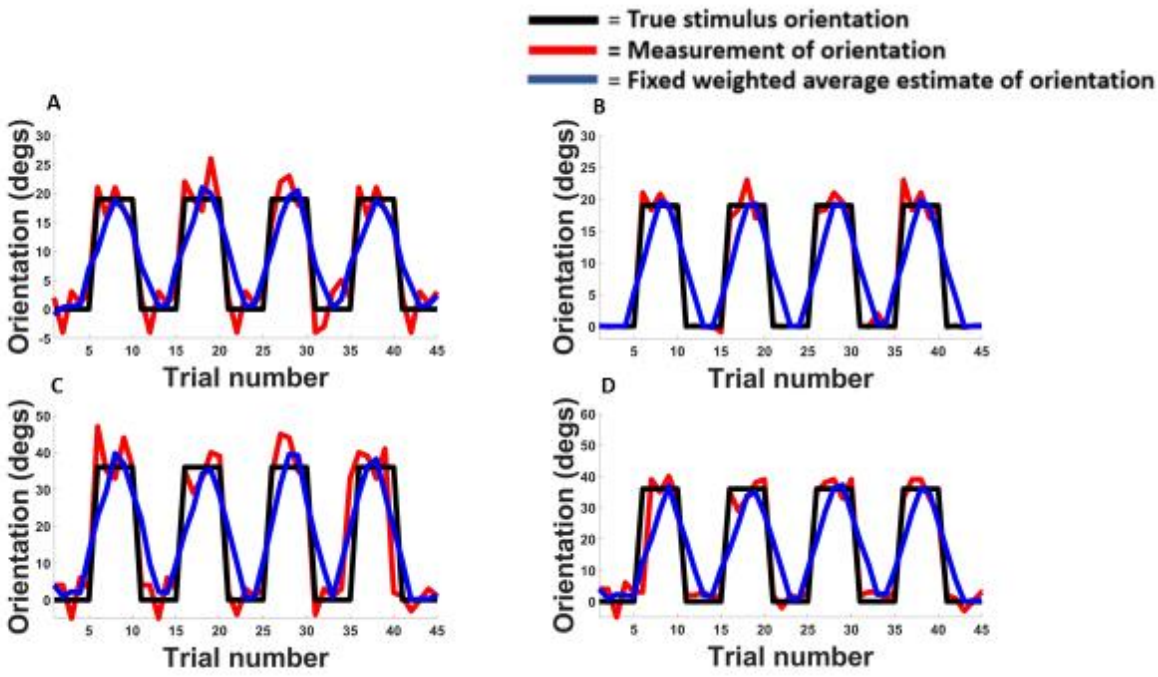


Figure 5. Simulated step response function for the fixed weighted average model with a weight of 0.7 on the current trial and 0.3 on the previous (n-1) trial. Note in all sub figures noise is smoothed over but estimates of changing values are lagged in time.

Kalman filter step response simulations

Here we provide a number of hypothetical step responses to our experimental conditions made by a Kalman filter under a number of different Kalman gains (figure 8) and also clarify exactly what constitutes the proximal and distal stimuli and their variance in our experiments in figure 7 (below).Our distal stimulus is the true stimulus orientation and distal variance the variance in those orientations over trials which we manipulate with large or smaller step sizes. The proximal stimulus is the participant's measurement of the stimulus orientation and proximal variance the variability of that measurement which we manipulate by presenting Gabors in higher or lower contrast.

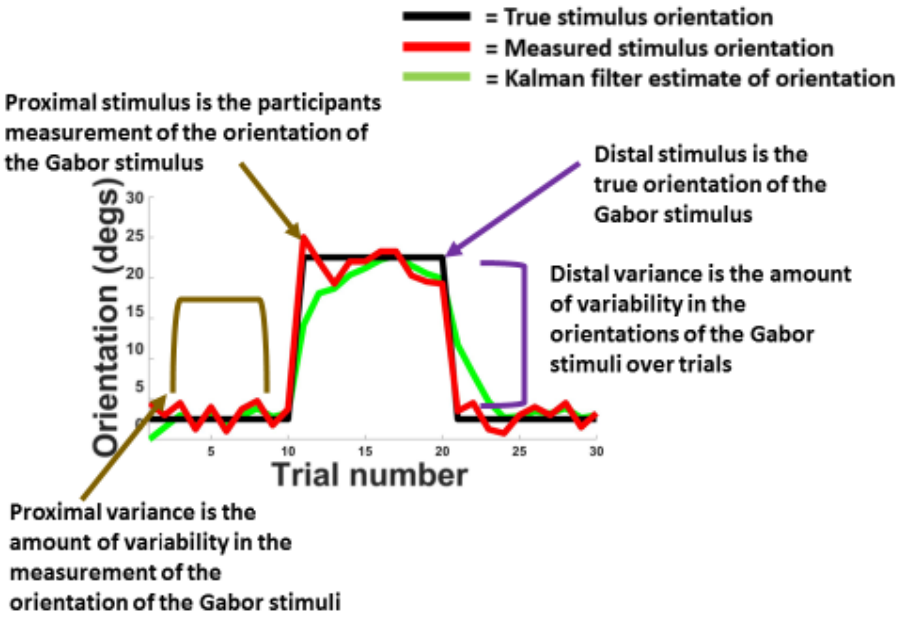


Figure 6. Simulation illustration of proximal and distal stimuli & variance in our experimental paradigm. For clarity, we illustrate exactly what constitutes proximal and distal variance in our simulations which also holds true for our actual experiments.

Figure 8, (below) shows a range of simulated Kalman filter step responses under a range of Kalman gains. Sub figure A, shows a simulated Kalman filter step response with to condition one which has low distal and high proximal **(**18.6° step/5% contrast). Here, due to the relatively high ratio of distal variance a high level of Kalman gain is formed and the Kalman filter corrects to the step in stimulus values almost instantly. Sub figure B shows simulated a Kalman filter step response to condition two which has low distal and lower proximal (18.6° step/20% contrast). In such conditions we might expect a lower Kalman gain to be calculated and in this case the step response should be slower to adapt to the step in values and slightly slower to reduce error to zero than in condition one. Sub Figure C show the Kalman filter step response to the high distal and low proximal condition three (35.7° step/5% contrast). In this case we would expect Kalman gain to be smaller due to the influence of

higher proximal variance and the step response in this condition to adapt and reach zero error more slowly sub figure D shows a simulated Kalman filter step response to condition four (35.7° step/5% contrast). Here we have a step that carries high distal variance and low proximal variance. In this case a higher Kalman gain should be calculated and confer a faster rate of correction than in condition three.

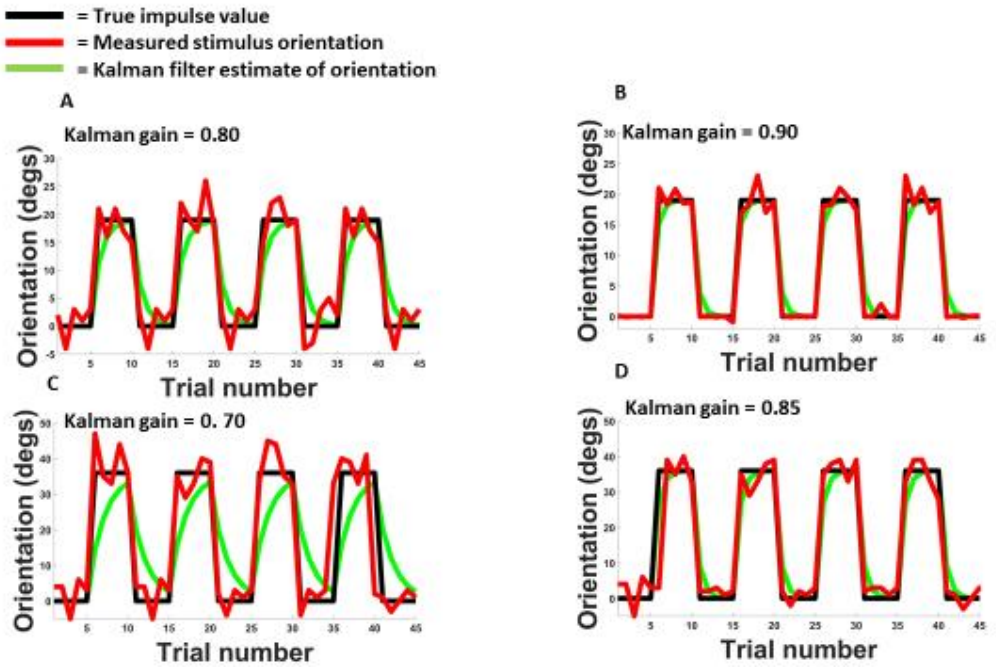


Figure 7. Simulated Kalman filter step responses to our four experimental conditions under a range of possible Kalman gains. Here we illustrate how Kalman gain determines how fast error is corrected. The closer Kalman gain is to one the faster error is corrected.

4.5 Results.

4.5.1 Proximal variance calibration experiment.

A Wilcoxon signed ranks test indicated that proximal variance ($deg^{2)}$ in the 5% contrast condition (Mdn=115.23) was significantly higher than in the 20% contrast (Mdn=9.68) Z=-3.18, p=0.00, r^2 = 0.77. Results indicate a very strong effect of contrast on response error variance. Large significant differences between contrast conditions also indicate that our experimental manipulation of contrast and our measurement paradigm were successful and this result is given extra validity by a very large effect size (Cohen, 1988) of contrast on error variability as shown below in figure 9.

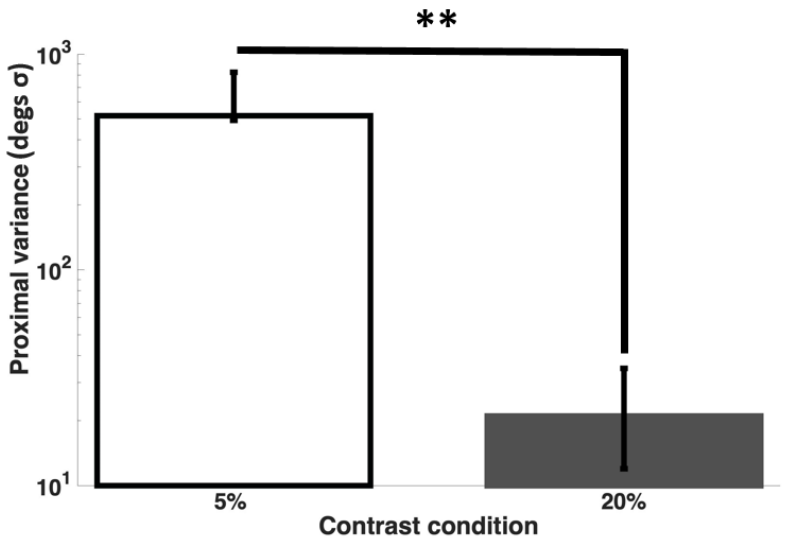


Figure 8. Group level median error variances taken from our all our individual psychometric fits. All participants recorded less variability and greater accuracy in the judgment of the stimulus orientation relative to the previous in the low proximal variance (20% contrast) condition in comparison to the high proximal variance (5% contrast) condition.

4.5.2 Individual level results.

In our individual analysis, we aimed to examine the general relationship between the level of proximal variance and the way error was corrected in participants' responses. We did this in two parts. In the first part, we simply graphically illustrate responses in our main experiment in a number of subjects who exhibited either high or low proximal variance levels in our proximal variance calibration experiment. In the second part, we split participants into the high and low proximal variance observer sets (5 highest and 5 lowest) and examine changes in the relationship between proximal variance and Kalman gain. We then provide step response plots for both high and low proximal observer sets as a further graphical illustration of this relationship.

Graphical illustrations of step response in individual high and low proximal variance observers.

This more qualitative type of analysis, led to some interesting observations consistent with the Kalman filters idea about the influence of measurement variability on the way errors are weighted. In participants in which proximal variance tended to be higher, there did seem to be a more pronounced lag to correct to the step in stimulus values in conditions 1 and 3 which presented Gabors in the high proximal variance, 5% contrast (see figures 12 and 13) than in participants who had lower levels of proximal variance (see figures 10 and 11). Also, when there was a large difference in proximal variance between 5% and 20% contrast conditions within subjects there also appeared to be a lag in response in the 5% contrast condition which is not so evident or non-existent in the 20% contrast condition. This effect is especially noticeable when examining the rate of correction to the trial immediately after the step in values which we highlight in the individual participant error plots we provide below. Figures 9 and 10 illustrate how two single low proximal variance observers (5% =46.24 & 13.67, 20% = 9.63 & 7.22) correct to the two different step sizes used in our main experiment. Here the rate of correction in both contrast conditions is almost immediate with some amount of overshoot to the change of stimulus value also recorded. Contrast this response pattern with the rate of error correction in the two individual high proximal variance observers (5%=152.27 & 341.67, 20% = 11.75 & 16.35) we illustrate in figures 11 and 12. In these participants, the rate at which error is corrected is slightly lagged in the 5% contrast condition and less overshot to the step in orientations. This difference in response, albeit rather small and inconclusive at an individual level is generally consistent with predictions made by the Kalman filter in which the rate of error correction, is in part, modulated by proximal variance.

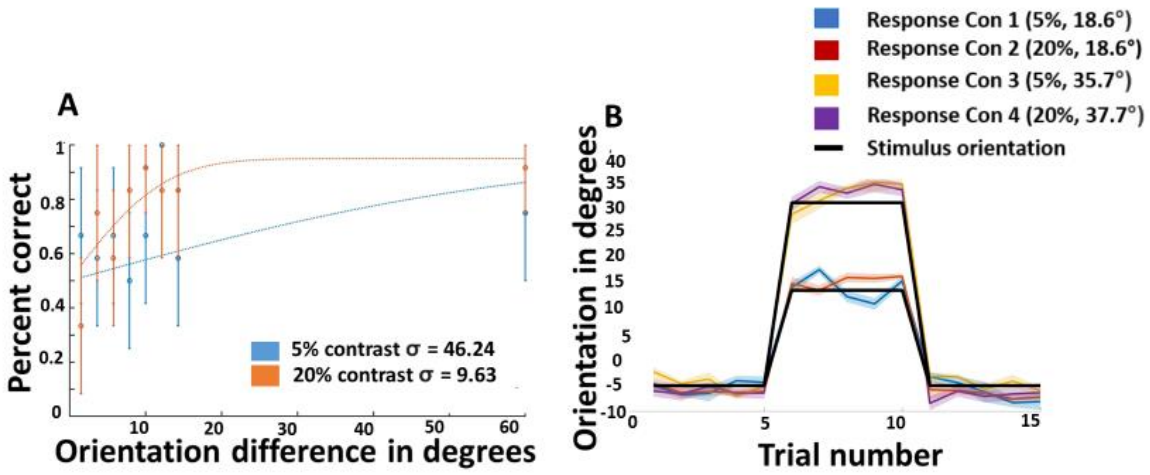


Figure 9. Example of step responses in a single low proximal variance observer. In the above figure we observe the rate in which a low proximal variance observer based on values recorded in the proximal variance calibration (A) corrects to the two different step sizes used in our main experiment (B). Here in both contrast conditions the correction to the step in stimulus values in almost immediate with some overshoot observed.

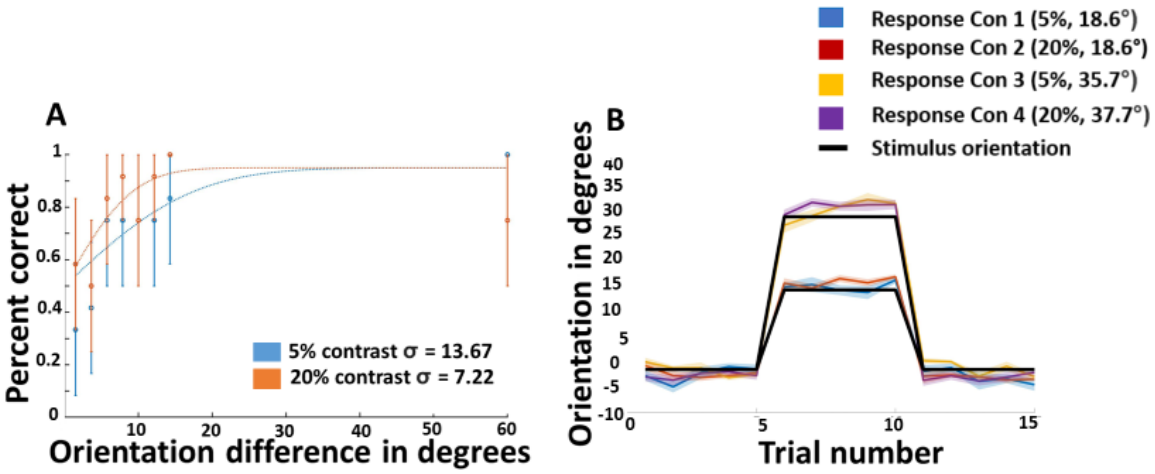


Figure 10. Example of step responses in a single low proximal variance observer. The above participant recorded very low proximal variance in both contrast conditions in the proximal variance calibration (subfigure A) and corrects instantly to the change in stimulus values in subfigure B.

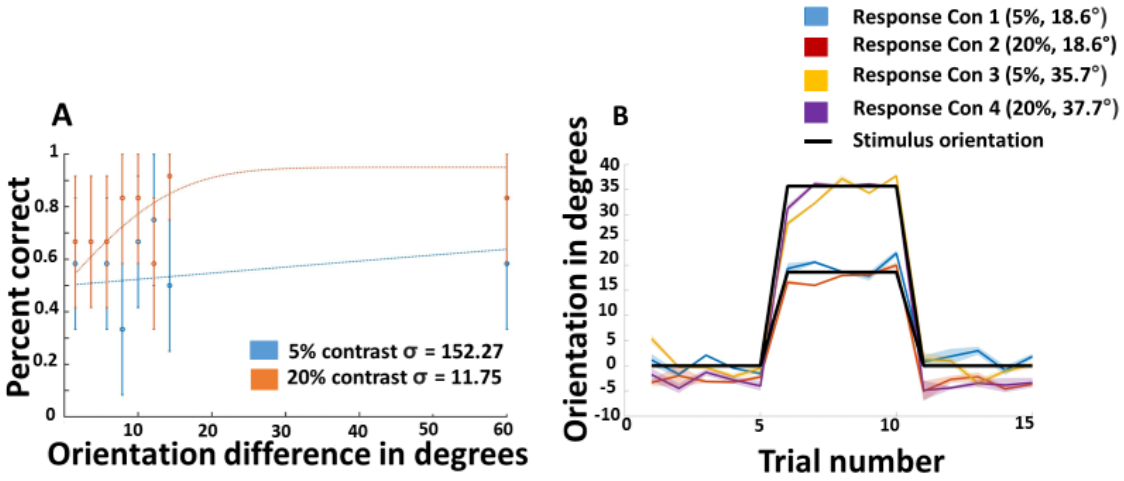


Figure 11. Example of step responses in a single high proximal variance observer. The above figure illustrates the lag in error correction (B) in a single subject who exhibited high levels of proximal variance in the 5% contrast in the proximal variance calibration (A) and average amounts in the 20% contrast. Note the lag in the rate of error correction in comparison to low proximal variance observers in figures 10 and 11.

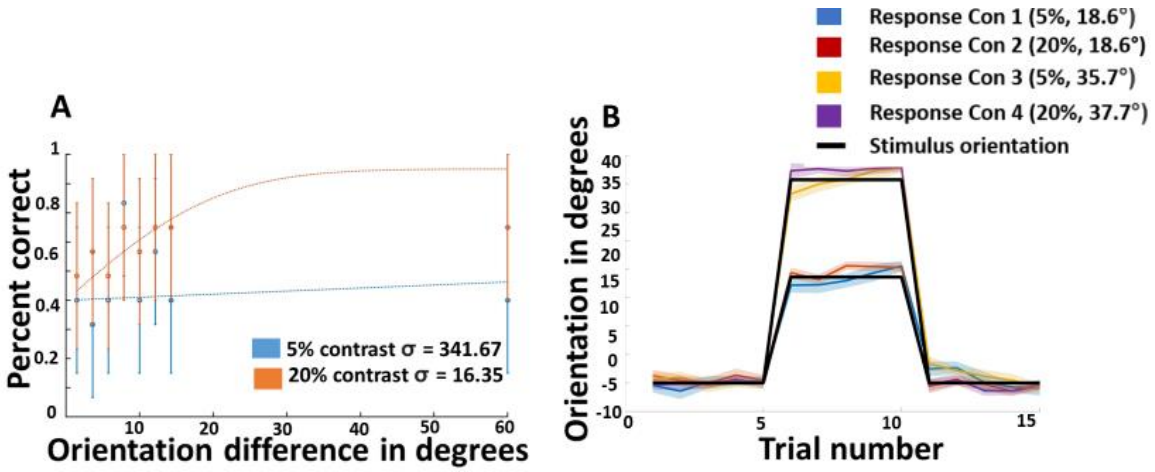


Figure 12. Example of step responses in a single high proximal variance observer. Here, we again observe a slight lag in error correction in a high proximal variance observer (A) in conditions 1 and 3. However, it is less pronounced than we might expect given the high level of proximal variance recorded in this participant.

To ensure our sets of low and high proximal observers had sufficient differences in proximal variance to allow a potentially detectable difference we only examined the correlation between proximal variance and Kalman gains in the 5% contrast condition (see figure 14).

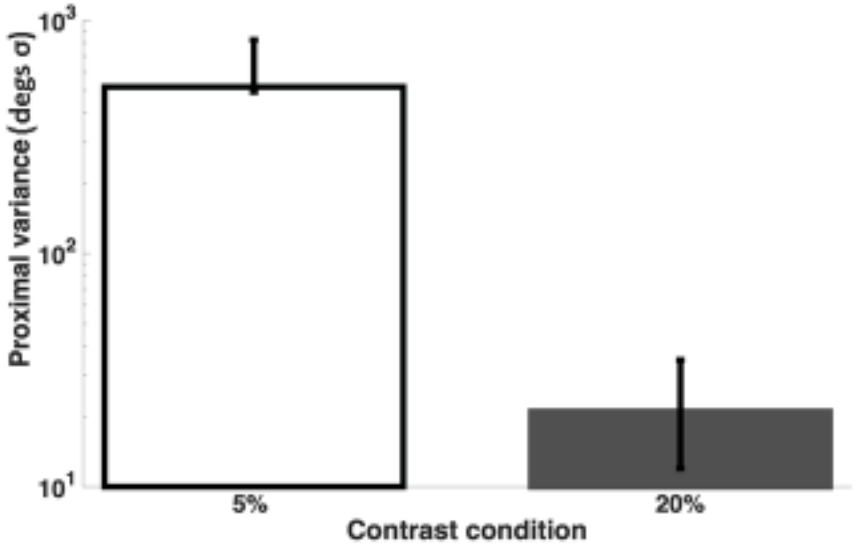


Figure 13. Median values in proximal variance between high and low proximal observers in the proximal variance calibration experiment in the 5% & 20% contrast conditions. Here we observe significant differences in median values in proximal variance between the top and bottom rated 5 participants in the 2AFC proximal variance calibration. Error bars are 25 and 75 percentiles.

4.5.3. Correlations between proximal variance and Kalman gain in low and high proximal observers

We analysed whether there were differences in the relationship between proximal variance and Kalman gain in high and low proximal observers in the 5% contrast condition by using a Spearman Rho correlation. In the high proximal observers in the 18.6 ° step there was a moderate but non-significant negative correlation between proximal variance and their corresponding model fitted Kalman gains (rs(4)=-0.56 *p*=>.05. In the high proximal observers in the 35.7 ° step there was a very similar result. In this case we again recorded a non-significant moderate negative relationship between proximal variance and their corresponding model fitted Kalman gain (rs (4) =-0.53 *p*=>.05. Interestingly, while we did not observe significant results there is a different pattern in the correlations between Kalman gain and proximal variance in the low proximal observers and high proximal observers. In the low proximal observers in the 18.6 ° step there was no relationship between proximal variance and model fitted Kalman gains with (rs(4)=0.00 p=<.05 while in the 35.7 ° step we recorded a moderate negative correlation with (rs(4)=-0.30 p=<.05. In summary, we observe moderate negative correlations between Kalman gains and proximal variance in the high proximal variance

observers in both step sizes indicating to some degree that when participant proximal variance goes up Kalman gain goes down.

Another aspect of the high versus low proximal observer's analysis we conducted was to simply compare step response plots for our sets of low and high proximal sets of participants. On this occasion, we included the 20% contrast condition as this aspect of the analysis is again designed to be a more graphical qualitative type comparison. Here, we observe differences in the rate of correction to the step in stimulus values between high and low proximal observers then in our previous individual plots of a similar nature which we illustrate in figures 15, 16, 17 & 18 below. In figure 15, we observe the step responses of all 5 low proximal observers. We can see a quite fast rate of correction in participants. In condition one (5%, 18.6° step), we do see a slight lag to but the effect is quite slight. In the lower proximal variance condition 2, (20%, 18.6 ° step) responses correct instantly and in fact overshoot by quite a large amount in a similar way we observed in some individual participants who exhibited low proximal variance that we illustrated previously in individual plots 10, 11, & 12. Responses in condition 3 (5%, 35.7 ° step), also exhibits some lag in correction to the step but it is quite small, while condition 4 (20%%, 35.7 ° step) responses correct instantly and in fact overshoot again, consistent with condition 2 which also had 20% contrast carrying low distal variance (see figure 15).

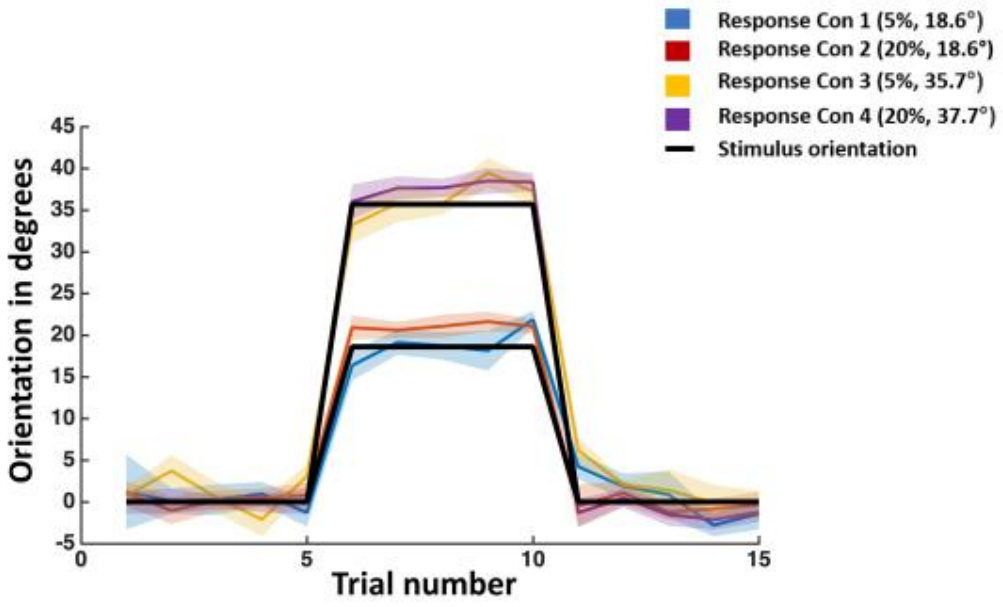


Figure 14. Group step response plot for the low proximal variance observer participant set. Here we note slightly slower correction in conditions with higher proximal variance (1 & 3), with some overshoot in lower proximal variance conditions (2 & 4).

A further graphical illustration that plots the same data displayed in figure 15 but shown in a different way is in figure 15 below. Here, we collapsed all step and stabilization phases into one step response comprising 15 trials (as they all are in the experiment). The aim of these plots is to focus on the

transition from step to stable phases in a number of important trials. The key trials are as follows; trial one, which is the first trial in the run of 15, trial 5 which is the trial just before the stimulus values steps before values actually step up on trial 6 and then stabilize post step on trial 7, trial 10 is the last trial in the post step stabilization phase, with trial 11 the first trial after the step changes back to its original pre step values with trial 15 the last trial in the run of 15. In sub figure A, we observe a lag in correction in condition one between trials 5 and 6. Sub figure B shows condition two, in which we see a fast correction to all presented orientation and some persistent overshoot between trials 6 and 10. Sub figure C illustrates responses in condition 3, which exhibit fast correction between transitions between step and stable phases and again some overshoot while lastly we have condition four in sub figure D. Interestingly in condition 4 responses also appear to show some slight lag between trials 5 and six but it is quite small and does not persist. In general, the picture is that in low variance proximal observers we do observe some slight lag but it is small and short in duration and a quite subtle effect.

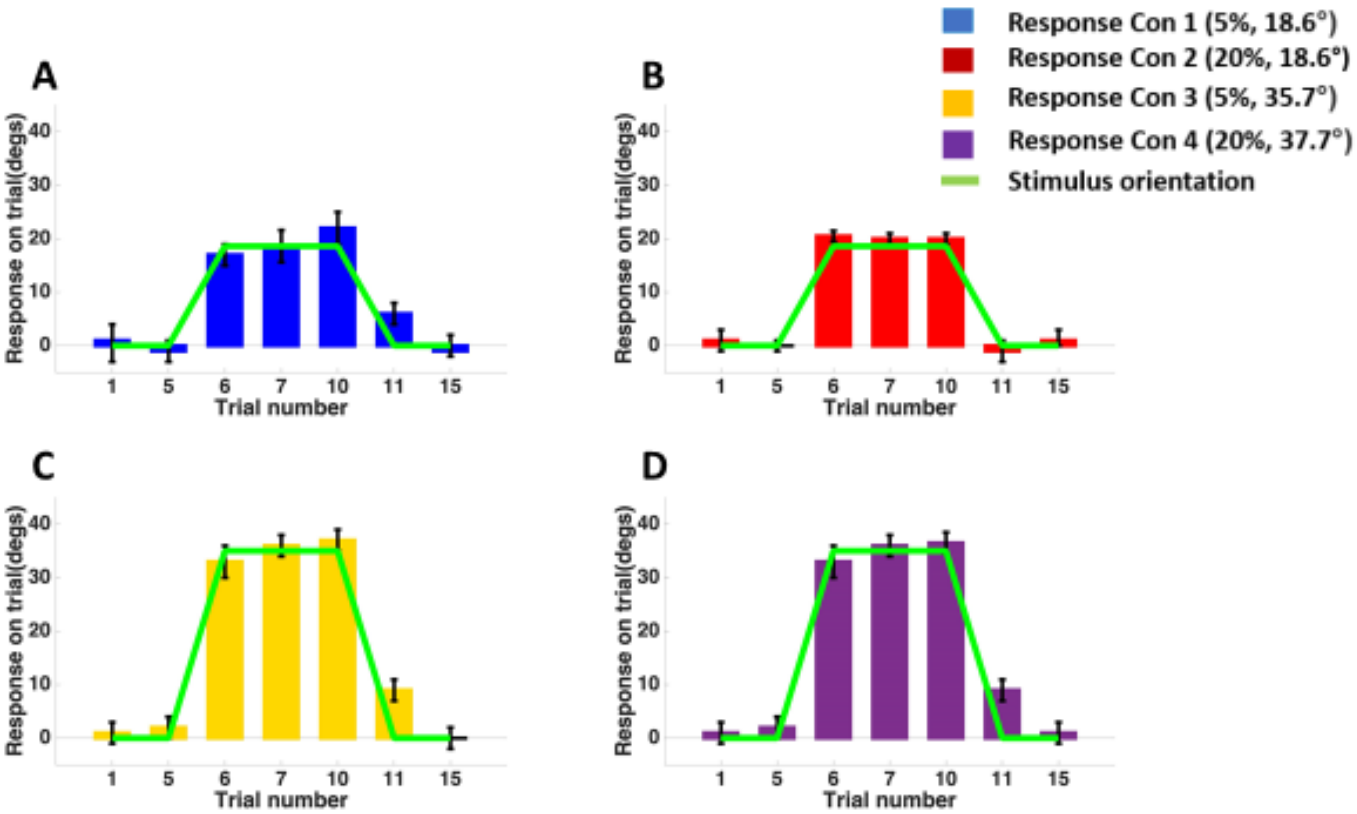


Figure 15. Bar graph showing the step response during key step to stabilization transitions trials in low proximal observers.

In figures 16 & 17, we illustrate the group step response functions for the high proximal variance observer group. In figure 17, we note a more pronounced lag in the way error is corrected in comparison to the low proximal observer participants. This is especially evident in condition one (5%,

18.6° step) in which the rate of correction is much lagged and does not correct until the end of the step phase. A smaller lag is also observed in condition three (5%, 35.7 ° step) in which error is nearly always present during the step phase although not as large as in conditions one. In conditions in which stimuli are presented at 20% (2 and 4) the overall outcome is that there seems to be a noisy response and some over shoot in both conditions and a faster correction to the step in stimulus values.

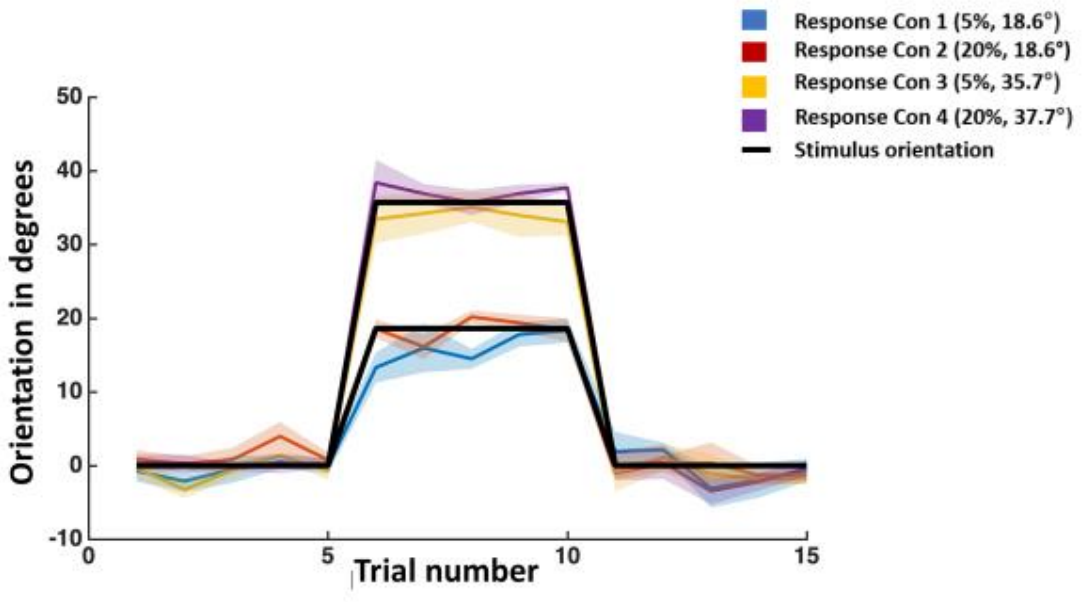


Figure 16. Group level step response plot for the high proximal variance observer participant set

Figure 17 again focusses on the transitions between step and stabilization trials in the high proximal observer's participants. Sub figure A, clearly shows the lag in error correction in condition one. The effect is especially clear on trial 6 and persists over time. As condition one has both the highest proximal variance and lowest distal variance meaning it would have the lowest Kalman gain of any condition this result strongly supports the Kalman filter and given we observe a much stronger result than in the same condition in the high proximal observers does suggest a role for proximal variance in error weighting. Sub figure B, shows condition two, here we have lower proximal variance than in condition 1 resulting in a faster correction over transitions from step to stable phases in trials 5 and 6. Sub figures C shows condition 3 which has 5% contrast (high proximal variance) again as in the other 5% contrast condition we observe a lag between trials 5 and 6 but admittedly it is small and short lasting. Lastly, sub figure D provides transition plots for condition 4. Here, we have a very large

overshoot on trial 6 and no lag visible at all, consistent with what we might expect for a low proximal/high distal variance stimuli in the Kalman filter and our simulation for a high Kalman gain

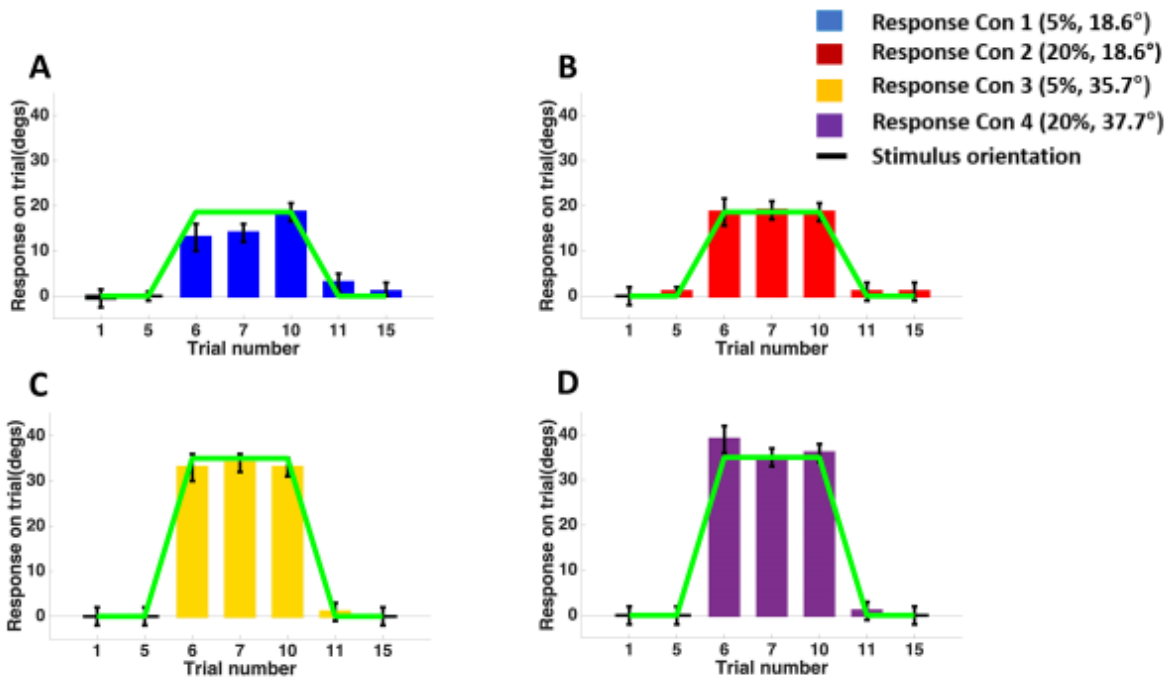


Figure 17. Bar graph showing the step response during key step to stabilization transitions trials-high proximal observers.

4.5.4. Main analysis.

The main analysis of the current chapter tested the existence of differences in model fitted Kalman gains between our experimental conditions in all participants. This analysis comprised a repeated measures ANOVA and post hoc tests in the shape of pairwise comparisons corrected for multiple comparisons and reported some interesting results consistent with the Kalman filter model of response error updating. However, before reporting our main analyses we again examine the general pattern of responses between our experimental conditions using our step response plots and bar graph summary of responses on key transition trials.

Figure 18(below) shows the group step response function for all conditions and all participants. Cleary visible and consistent with the Kalman filter we have a pattern of results that could be considered to be the levels of Kalman gain we might generally expect for each condition. While conditions 2 and 3 might provide quite close Kalman gains figures, as condition one has high proximal variance and low distal variance, we would definitely expect this condition to have the lowest Kalman gain and condition four with low proximal and high distal to have the highest Kalman gain. This means that condition one should be slowest to correct and condition four to be fastest to

correct. This is exactly what we see and supports the adaptive error weighting strategy outlined in predictive coding and the Kalman filter a result again well illustrated in our key transitions plots (see figures 19 and 20 below).

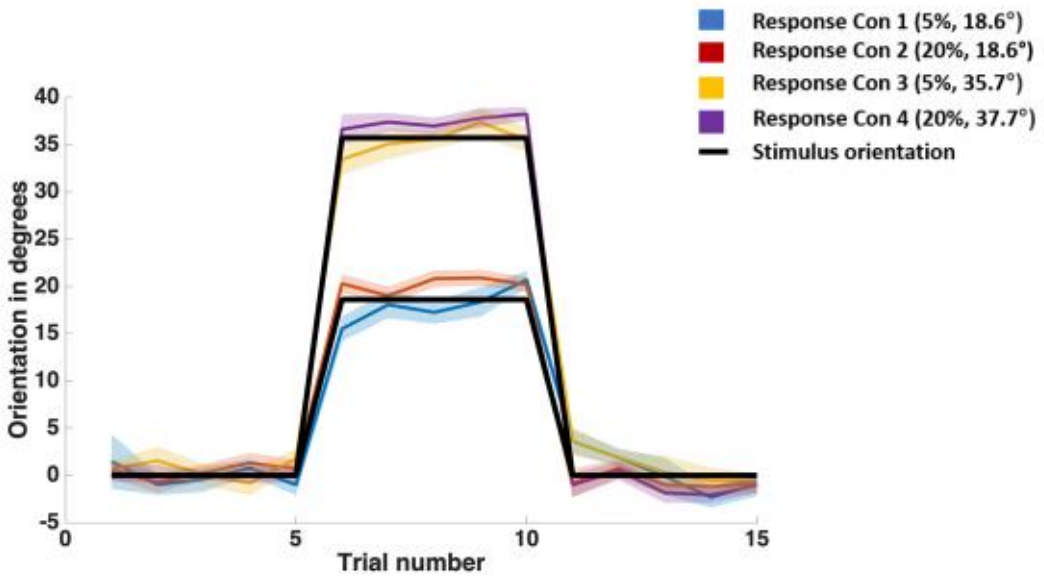


Figure 18. Group step response plot for all participants in all conditions. Note the lag in correction and absence of overshoot in the high proximal variance conditions 1 and 3 in comparison to the low proximal variance conditions 2 and 4.

Figure 19, plots participant's responses on key transitions and complements the pattern of results we see in figure 19 above. Condition one, exhibits the slowest correction to the step in values on trials 5 and 6, and interestingly shows a small amount of overshot on trial 10. The step response in condition 2, corrects instantly to the change in values on trial 6 and actually remains slightly above the true stimulus values consistent with previous individual results reporting overshoot in conditions with high distal variance (20% contrast condition). Condition 3, representing the second high proximal variance condition (5% contrast) with condition 1, also exhibits a slight lag between trials 5 and 6 but is smaller than that in condition 1, while condition 4 with low proximal but high distal variance shows an instant correction to steps in orientations with quite a large and persistent overshoot.

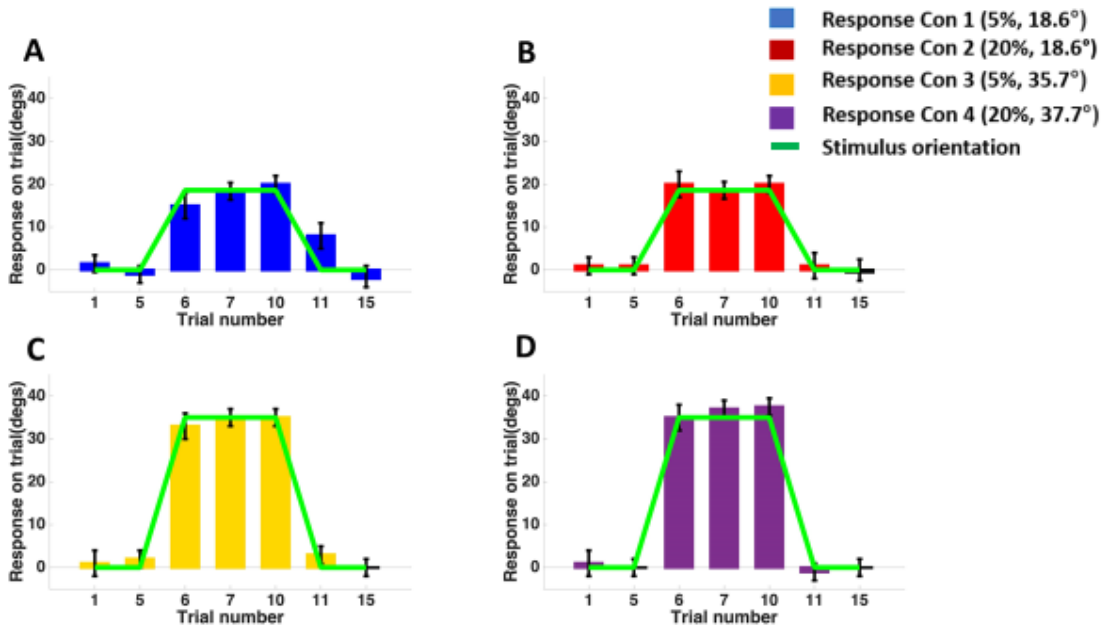


Figure 19. Bar graph showing the step response to selected orientations for all participants. In this figure we observe pre step responses on trials 1 and 5 at around 0 in all conditions. Conditions 2, & 4 are relatively fast to correct to the step in values at trial 6 and with condition 4 displaying some initial overshoot. However, the above plot show the lag in correction to the step in values at trial 6 in condition one and the small lag in condition 3 in condition one and a slight lag in condition 3.

Main effects.

A repeated measures ANOVA tested the effects of our four proximal/distal variance conditions on model fitted Kalman gains. Analysis reported a significant effect of condition on model fitted Kalman gain $F(1, 10) = 5.969$, $p=0.002$. Pairwise comparisons (Bonferroni adjusted) were conducted to see where differences in mean model fitted Kalman gains existed between conditions. Results reported that mean model fitted Kalman gains in condition 1 (M=0.78, SD=0.63) were significantly different from condition 2 (M=1.04, SD=0.34), $p=0.010$ and condition 4 (M=1.01, SD=0.27), $p=0.012$ but not condition 3 (M=0.85, SD=0.56). Comparisons also reported that mean model fitted Kalman gains in condition 2 were significantly different from condition 1, $p=0.010$ and condition 3, $p=0.016$ but not condition 4. They also reported that mean model fitted Kalman gains in condition 3 were significantly different from condition 2, $p=0.016$ and condition 4, $p=0.015$ but not condition 1. Finally, our last pairwise comparison also reported that mean model fitted Kalman gains in condition 4 were significantly different from condition 1, $p=0.012$ and condition 3, $p=0.015$ but not 2. Overall, results

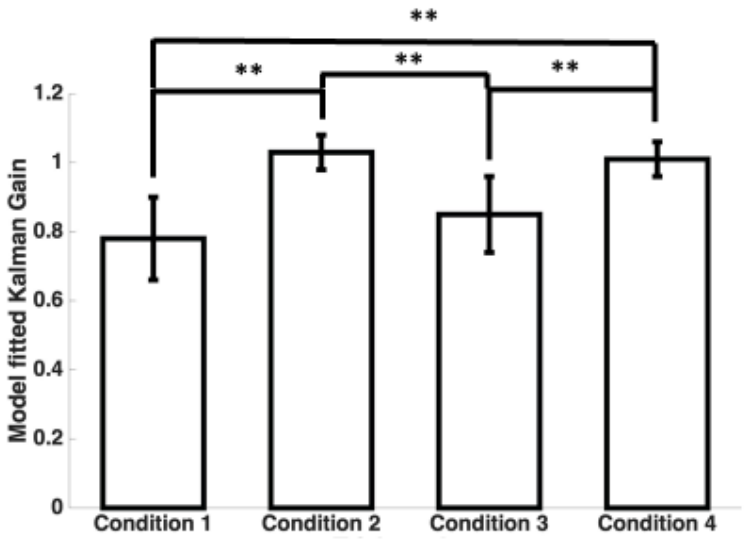


Figure 20 Bar graph showing the pattern of significant differences in Kalman between conditions. Here we summarise the pattern of significant differences in model fitted Kalman gains between our four conditions. Here, we show that the level of Kalman gain in condition one was significantly different from conditions two and three. The level of Kalman gain in condition two was significantly different to that in condition one and there. The level of Kalman gain in condition three was significantly different from conditions one and four and finally the level of Kalman gain in condition four was significantly different from conditions one and three

4.6 Discussion

The current chapter aimed to build on and extend findings from chapter 3 in regard to testing the adaptive weighting of prediction errors outlined in predictive coding against the fixed weighting of stimulus values implied in perceptual averaging. While results from chapter 3 supported predictive coding our use of serial dependence, while effective for the aims of the experiment was not an ideal behavioural measure of how error on current trials is reduced directly. This was because serial dependence is more of measure of the effects of past values on perception as opposed concentrating on current values. Furthermore, serial dependence experiments usually present stimuli that changes on every trial making it hard to ascertain the time frames over which error is reduced to zero or which factors might affect such reduction. To provide a means for us to better assess such issues we used a stimulus design inspired by the use of step response functions in control theory. By presenting stimulus orientations that remained stable for a fixed time period and then stepping in values for the same amount of time under different experimental condition based on Kalman filter theory we were able to formulate clear experimental hypotheses to test the predictive coding account of error correction against the perceptual averaging account of perceptual updating. Results showed that the way participants adapted to the step in stimulus orientations were consistent with the predictive coding account of error weighting and as in chapter 3 again well predicted by Kalman filter theory.

Furthermore, aside from what can be considered a novel overall finding our methods also provide techniques that could be used in future studies that examine predictive coding. However, the study does raise some interesting issues that require further explanation. Not least, the consistent overshoot we observed especially in conditions in which stimuli were clearly visible. We now discuss our study beginning with the validity of the two keys aspects of our methods-the proximal variance calibration and the step response stimulus design.

General validity of experimental methods.

An important aspect of our analyses at both an individual and group level was the influence of proximal variance on the way participants corrected to steps in stimulus values. For us to quantify this relationship we needed an accurate measure of this source of variance. The level of distal variance was a known quantity as we knew our step sizes but the level of proximal variance is very much participant specific. To provide such a measure, we used the same proximal variance calibration experiment we utilised in chapter 3. Results recorded large and highly significant differences in median values between 5% and 20% contrast conditions, indicating that contrast had a substantial effect on the perception of our stimuli further supported by a large effect size. Such a result is in line with previous findings indicating factors such as blur (Kayargadde & Martens, 1996), luminance (Waugh & Levi, 1993) and most importantly contrast (Waugh & Levi, 1993) affect perceptual measurement variability. Our results also provide a direct replication of our findings from chapter 3 in a totally new participant set and indicate our proximal variance calibration paradigm is a reliable means to ascertain levels of proximal variance both within and crucially across participants. There are though, however, questions of our interpretation of these results in terms of purely visual proximal variability.

One potential problem in the interpretation of findings from our proximal variance calibration experiment as purely visual variability stems from the way we presented our stimuli. Our experimental design used a 2 alternative forced choice paradigm. We presented an initial Gabor at a certain orientation and then presented a noise screen for 500ms, then a fixation cross for 250ms, then a second noise screen and following that the second Gabor orientation about which participant's had to make a judgment about. This meant that there was a time gap between from the first Gabor orientation and the second orientation and subsequent judgment of approximately 3 seconds. This delay raises an issue because there is inevitably some level of working memory required by participants to make the judgment of orientation as the task involves a judgment involving a past stimuli orientation no longer on the screen. As with all neural computations, working memory is subject to some level of variability both within (Fougnie, Suchow, & Alvarez, 2012) and across participants (Fougnie et al., 2012) which may be at quite different levels for the same participant's perceptual variability levels. Due to the use of working memory and the input of variability from this

cognitive system it is open to question whether it is entirely accurate to consider our data for proximal variance as only arising from perceptual systems. Of course, one could argue that there is not really a distinction between perceptual and visual working systems and if there is where does perception end and working memory begin. This is perhaps something to consider, not only here, but perhaps in all 2 alternative forced choice type experiments. However, while recognising this limitation we contend that as our proximal variance calibration experiment involved the same time delays in stimulus presentation as our main experiment we could still make a valid comparison for 'total' proximal variability and its effect on adaptation rate to our changes in stimuli in our main experiment at an individual level and group regardless of its source.

The second key aspect of methods for discussion here is the novel use of a step response type experimental design in conjunction with Fischer & Whitney's (2014) basic design in purely visual terms. On the whole we found this to be a successful experimental paradigm. While the step design has been effectively applied in visou motor research to examine how errors reduce over time (Baddeley et al., 2003; Burge et al., 2008) it had not been widely used in vision outside of the smooth pursuit literature (Brostek et al., 2017), representing much different types of experiment using smaller time scales. Given that visuomotor studies deal with two sources of variance (motor and visual system) meaning that more uncertainty is present in estimates it may have been possible that as orientations repeated a number of times it would have been too easy a stimulus to look at the way error corrects over time in a meaningful way and participants would simply have responded at the current orientation in all conditions. This was not the case and our experimental design was able to show how proximal and distal variance modulated error correction extremely well. However, there are some minor critiques in its use in the current chapter specifically and in more general use. In the current chapter we use the Kalman filter to model and explain participant step responses. Technically, the optimal use of the Kalman filter requires stimulus values to have a normal distribution which the step response does not have. This though it could be contended is not really an important in the current chapter as we were really only using the Kalman gain to model general response patterns and not to estimate actual values. Also, a common critique of the use of step response functions in control systems applicable to human experimentation is that they are an artificial stimuli. In the real world stimuli or the signals they emit do not behave in such a manner but nonetheless they are excellent for comparing the outputs of different systems on an equal footing. Overall, the step experimental design worked well and provided some excellent results in our main experiment.

The key manipulation in the current chapter was the way we altered proximal and distal variance across our conditions (condition 1, 18.6 °step/5% contrast, condition 2 18.6° step/ 20% contrast, condition 3, and 35.7 ° step/5% contrast and condition four 35.7 ° step 20%) While we could be certain that we had increased the level of distal variance in conditions 3 and 4 which had a 35.7° step which will inevitably carry more variability than the smaller 18.6 degree step we had to be certain that

we had significantly different levels of proximal variance between the 5 and 20% contrast conditions to validate our manipulations. While we had achieved such results using identical methods in chapter 3, in this chapter we had a totally new participant set and as we had observed quite large inter individual differences in proximal variance in chapter 3 we could not be certain about what we might observe on this occasion. Furthermore, as we intended to look in greater depth at the role of individual differences in proximal variance and the way error was corrected over time.

Our individual comparisons of proximal variance psychometric fits from our calibration procedure and the way participants responded to changes in stimulus orientations had two purposes. The first was of course to compare the way distal and proximal variance affected how participants' integrated information and corrected error but the second was to build on findings from experiment three which indicated that when participant's exhibited a more noisy response they appeared to become more 'serially dependent' with responses closer to previous values. While chapter 3's main experiment tested different participants to its proximal variance experiment, here we tested the same participants. This method allowed a clearer individual level comparison.

Individual comparisons of proximal variance psychometric fits from our proximal variance calibration experiment and the way participants corrected to steps in stimulus orientations in our main experiment provided some interesting if inconclusive results. In participants who recorded low levels of proximal variance there appeared to be an almost instant correction to the step in stimulus orientations across all four experimental conditions (a result consistent with a Kalman filter like responses (Harris & Wolpert, 1998; Todorov, 2004) as opposed to a fixed weighted average response (Choo & Franconeri, 2010; Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Fischer & Whitney, 2014). However, without a comparison to low proximal observers such a response pattern may simply be attributed to inter-individual noise in responses as we did observe a large range of response variability across participants. Comparison with a limited number of high proximal observers, did though, again give credence to an adaptive, Kalman filter explanation of participant's correction to the step in stimulus orientations again consistent with findings from visou motor adaptation research (Burge, Girshick, & Banks, 2010; Wolpert, 2007; Wolpert & Flanagan, 2001). In high proximal observers, there did appear to be more of a lag in correction to the step in stimulus orientations in all conditions. However, the lag seemed to be also present in conditions which presented Gabors at 20% contrast which we might not have expected to find. Furthermore, the level of difference in correction between conditions in individual participants was small and while this aspect of our analysis was only intended as a qualitative comparison and while still worthwhile lacked statistical comparisons making interpretation difficult.

Comparison of high and low proximal observer sets

The next analysis in our study provided a more statistically based analysis of the effects of proximal variance and the rate of error correction. In this analysis, we split participants into two groups (high and low proximal variance observer sets), quantified model fitted Kalman gain as a measure of the rate of error correction and formulated experimental hypotheses to test the strength of the relationship between proximal variance and Kalman gain. Here, hypotheses were based on a theoretical inverse correlational relationship between proximal variance and Kalman gain. In the Kalman filter, if distal variance is relatively constant, when proximal variance goes up Kalman gain should come down and when proximal variance goes down, Kalman should go up. This means that if a set of participants have high proximal variance we could have expected to record lower model fitted Kalman gains than in participant sets with low proximal variance. Importantly, due to the small range of proximal variance across the board in participant's in the 20% contrast condition, we only analysed the data in conditions which presented Gabors at 5% contrast as with such a small range observing any differences in correlation between our two participant's sets in conditions using 20% contrast (3&4) was highly unlikely (see results figure 14). We hypothesized that we would find a higher negative correlation between Kalman gain and proximal variance in high proximal observers than in low proximal observers which is somewhat open to debate.

The first aspect of our results in our correlational analysis to note is that all of our correlations between proximal variance and Kalman gain were non-significant. However, we contend that results still provided support for the existence of an inverse relationship between proximal variance and the way errors are corrected in line with previous literature in the field (Burge et al., 2008; Denève et al., 2007; Friston, 2010; Harris & Wolpert, 1998; Knill & Pouget, 2004). In our high proximal observers in condition 1 we recorded a moderate negative correlation of -0.56 between proximal variance and Kalman gain, while in condition 3 we recorded a similar negative correlation of -0.53. Contrast this with our correlations from our low proximal variance observer set who recorded a correlation of exactly 0 in condition one and in fact a positive correlation of 0.30 in condition 3 suggesting that at least to some degree proximal variance did exert an effect on the way participant's adapted to the step change in stimulus orientations. This interpretation is given extra validity if we further consider our response plots and bar graphs (see results figures 15, 16, 17 & 18) and the relationship between correlation coefficients and *p* values and the way Kalman gain is calculated.

In our analyses we only tested a small number of participants in each of our low and high proximal observer sets. If we focus on the correlation coefficients (which is not affected by the sample size) between Kalman gain and proximal variance as opposed to *p* values (which are affected) and consider the way Kalman gain is calculated (see chapter 2) then we did record perhaps recorded the upper end of any correlation we might have observed. This is because Kalman gain is not only calculated based

on proximal variance but also includes prediction variance mainly created by distal variance (Zarchan & Musoff, 2000). By only looking at proximal variance, we only analysed one part of the causes of any changes in Kalman gain meaning that the correlation between Kalman gain and proximal variance can never reach one. Indeed, in the calculation for Kalman gain, prediction variance is the dominant factor, as this is in fact what the model is trying to 'keep' while proximal variance is considered noise and to be discarded. If we consider these factors, a correlation of around 0.5 in our high proximal observer set between proximal variance and Kalman gain, may in fact, given our small sample size been a very good result and give solid support to the Kalman filter model of error correction. However, to really test the Kalman filter versus the fixed weighted average account of error correction we need to take into account both distal and proximal variance and their relationship to Kalman gain analysis which formed the basis of our main experiment. However, we do note that correlations are a noisy statistical measure and perhaps providing more detailed and dedicated experiments that look at the role of proximal variance would in which proximal variance is manipulated in a number of ways and with more participants would provide a fuller picture due to the limitations in our experiment.

Main experimental findings

The aim of our main experiment and analyses was to fully test the adaptive account of error correction posited in the Kalman filter against the fixed model of error correction outlined in the fixed weighted average models. In contrast to our previously discussed analyses here, we included all participants and tested Kalman gain directly as a proxy measure of error correction against all four of our experimental conditions. As our experimental conditions manipulated both proximal and distal variance in a number of permutations we were not limited to looking just at the effects of proximal variance on error correction. Our experimental manipulations built on our manipulations of proximal and distal variance in chapter 3 and also ideas from Burge, Ernst and Banks (2008) who also tested four combinations of what can be considered proximal and distal variance. Again, our theoretical framework provided clear and testable experimental hypotheses. Specially, according to the Kalman filter that there would be differences in error correction caused by our manipulation between experimental conditions and according to the fixed weighted average model responses should not adapt. Our analysis reported results strongly in line with the Kalman filter with significant differences in the Kalman gain, indicating that error was adaptively corrected between conditions commensurate the level of Kalman gain.

Inspection of our step response plot for all participants (results, figure 19) gives an indication of the differences in the pattern of responses and error correction between conditions. In condition one (5% contrast) we observed a lag in response to the step in stimulus orientations and the time taken to correct error while in condition two while in the condition 2 (20% contrast) which also had an 18.6 degree step the lag is not present. In condition 2, responses to the step correct instantly and indeed

include some overshoot which persists until the end of the stabilization phase. In conditions which had a 35.7 degree step (3&4) we observed a similar pattern of error correction. Responses in the 5% contrast condition 3 also appear to lag behind the step in stimulus while in the 20% contrast condition 4 again we see an instant correction and again some persistent over shoot. Results from all four conditions are consistent with Kalman gain figures for the commensurate conditions and give general support for the Kalman filter model of error correction as opposed to the fixed weighted average account. However, before discussing our statistical analysis of Kalman gain we discuss a prevalent factor in many of our response plots namely the overshoot in responses seen to some extent in all conditions in our response plots and bar graphs at the group level (see figures 19 & 20)

Overshoot is often observed in many signal processing applications which test the response properties of a system using a step response function. One common reason for overshoot, especially in filters which employ a model of the stimulus behaviour, such as the Kalman filter, is that the model of the stimulus given to the system is incorrect. One function of the stimulus model is to constrain the next estimate within a set of known parameters. If these parameters are too 'wide' then estimates can increase over the true value of the stimulus. This could be the same cause of the overshoot observed in our participant responses, as unlike a signal processing system, we cannot simply programme in the model of the distal stimulus and it must be learnt. While it may be considered that the step in stimulus orientations is a simple model of behaviour to learn, in nature stimuli do not generally occur in steps and tend to change in a more correlated way over time which may make learning more difficult. This explanation is given validity by results from chapter 3's main experiment two, in which participant's did appear to learn the random walk (Pearson, 1905) of stimuli in some instances, possibly due to the fact that a random walk is a model closer to how stimuli generally behave in the world (Einstein & Cowper, 1926).

Analysis of Kalman gain differences between conditions recorded significant differences between a number of conditions (see results figure 21) in a manner almost entirely captured by our Kalman filter model predictions and simulations with responses modulated by the level of proximal and distal variance. Results reported significant differences in Kalman gain between condition 1 (0.78) and conditions 2 (1.03) and 4 (1.01), conditions 2 and conditions 1 and 3 (0.85), condition 3 and conditions 2 and 4 and finally between condition 4 and conditions 1 and 3. Furthermore not only did we record significant differences between conditions as predicted we also recorded the general ordering of Kalman gain level predicted by Kalman filter theory. Specifically, we predicted that condition one would have the lowest Kalman gain, with condition 3 next and 2 and 4 having the highest levels. This results provided very strong support for the Kalman filter model of correction over the fixed weighted average account and are also consistent with results from Burge, Ernst and Banks (2008) and a host of visou motor literature that posits an adaptive account of error correction (Denève et al., 2007; Harris & Wolpert, 1998; Knill, 2007). However, while we consider our

experimental paradigm to be very successful we also note that the study has its limitations especially the interpretation our findings in purely visual terms.

Previously we discussed the limitations of our interpretation of our proximal variance experimental data purely in visual terms due to the possible input of variability from working memory. A similar critique may also apply to our experimental design in our main experiment and our interpretations of our findings which may also have been affected by variability from other sources. One source of additional variability in responses in our main experiment could have arisen from the motor system. Unlike our proximal variance calibration experiment which used a key press to signal a judgment of orientation, in the main experiment participants signalled their judgment of the orientation using a mouse which is obviously controlled by the hand. It is well accepted that the motor system is highly variable (Faisal, Selen, & Wolpert, 2008). Furthermore, as we have discussed in depth it is also well reported that the motor responses exhibit Kalman filter like response properties in its responses to reconcile noise (Denève et al., 2008; Kwon, Tadin, & Knill, 2015; Wolpert & Flanagan, 2001) which may have exaggerated the level of bias to previous stimulus values we observed. Indeed, due to the level of noise in the motor system it is likely no coincidence that visou motor literature has pioneered the use of filter models to explain how sensory systems resolve variability. When we factor in variability from the motor system and from working memory we can logically ascertain that some of the bias we observed may have been the result if integration processes designed to remove variability in non visual systems. However, while the effects of filtering non visual variability may be a potential confound perhaps a more important issue comes from recent work that has indicated that the observed bias found in a number of types of stimuli (Corbett, Fischer, & Whitney, 2011; Fischer & Whitney, 2014; Kiyonaga, Scimeca, Bliss, & Whitney, 2017; Kramer et al., 2013; Liberman, Fischer, & Whitney, 2014; Taubert et al., 2016) may not be perceptual at all and arise from a post perceptual decision making process.

Conclusion

In conclusion, there are number of key findings from the current chapter that provide further understanding of how visual systems deal with uncertain information. One is that the correction of perceptual error is adaptive and adaptive and modulated by the variability of sensory measurements (proximal variance) and the behaviour of the stimuli of interest (distal variance). When proximal variance was higher relative to distal variance participants estimates corrected faster when stimulus values changed and when distal variance was higher relative to proximal variance estimates corrected slower to changes in stimulus values. This result is consistent with findings from chapter 3 and with findings from Bayesian visou-motor literature (Berniker & Kording, 2011; Burge et al., 2008; Denève et al., 2007; Wei & Koerding, 2010) and shows the Kalman filter can also be applied well in vision. The way participants integrate information over time exhibits high individual variability. Some of this

variability can be explained by the different levels of proximal variability in participant's observations but not all. Interestingly, again we found that temporal integration strongly favours the current stimulus input over previous information. Although, we observed significant differences in Kalman gain which acted as our proxy measure for error correction, Kalman gains did not drop below the 0.7 level in any condition indicating a much higher effect of the current stimulus value on estimates. To conclude results provide support for the predictive coding account of temporal integration and the use of ideas such as the Kalman filter and stimulus step designs to study how the visual system deals with uncertainty.

Chapter 5. Predictive coding as a dynamic process: the role of conditional relationships and sequential information in predictions and behaviour.

5.1 Abstract.

Predictive coding suggests that the brain extracts the temporal regularities present in the environment over multiple timescales and levels of complexity to optimize the level of prediction error in its predictions. One issue with this idea is that in the world there are many different types of temporal regularities of differing complexities and it is unclear whether the brain applies more complex probabilistic sequential regularities when making predictions. Previous predictive coding studies have commonly used cue type relationships to induce predictions or have biased predictions by presenting one stimuli much more often reminiscent of priming studies. Furthermore, studies that have used conditional sequences of stimuli have often shown that either they do not improve behaviours or that people have represented sequences in a non optimal way. Here, to provide further analysis of these issues we used an experimental design that presented a sequence of stimulus positions and omission trials in which stimulus changed in such a way that made certain changes more or less predictable based on two direction switching Markov chain transition probability matrices that manipulated predictability between screen positions and omission trials. Participants were asked to respond as quickly as possible to the stimulus screen position and that in some cases the stimulus would be very hard to see (omission trials) but they should press where they think it should be and recorded reaction times on visible and percentages pressed relative to previous trials on omission trials. We formulated two theories about what we might find. The first, motivated by predictive coding states that participants should always advantage the most probable screen position. If this was the case then reaction times would always be faster on predictable trials and when faced with an omissions trials they would press to signal the most predictable screen position relative to the previous trial. Our second theory was based on findings from priming studies that have indicated that repetition of consecutive trials is a more reliable determining factor in behaviours than sequential probabilities. In this case reactions times and presses signalling screen position on omission trials would be independent of predictability. Results reported a significant effect of predictability on reaction times and a significant interaction between reaction time and the directions of our markov chain in support of predictive coding. However, analysis of omission trials led to a somewhat confusing picture. Analysis suggested that participants had learnt the stimulus sequences but did not apply them in the manner suggested by either priming or predictive coding. Results suggest that participant's responses were based on a combination of a non optimal strategy termed probability matching and not necessarily applying any explicit prediction on omission trails. Overall, the main conclusion is that we can use sequential transition information that is more complex than simple cues or frequencies to improve behaviour but the relationship between reaction times, prediction error and decision making is unclear and warrants further investigation.

5.2 Introduction.

Visual information received at the cortex contains not only variability from signals emitting from stimuli in the world that are important for behaviour but also unwanted variability often termed noise (Faisal et al., 2008). External factors such as poor viewing conditions and anatomical factors such as saccades (Melcher, 2011) disrupt information bearing light signals coming from stimuli in the world with neuronal noise adding further variability making interpretation potentially problematic (Kayargadde & Martens, 1996b; Tuzlukov, 2002; Wolpert & Flanagan, 2001). Exactly how the visual system counters the effects of unwanted variability is a matter of keen theoretical debate. Two strategies implicated in helping to interpret uncertain signals are to average past and current information over time (Bauer, 2017; Burge et al., 2008; Corbett & Melcher, 2014; Corbett et al., 2011; St. John-Saaltink et al., 2016) and adaptively apply prior knowledge to help distinguish stimulus variability from unwanted variability or noise (Denève, Duhamel, & Pouget, 2007; Friston, 2010; Knill & Pouget, 2007). Previously we have shown evidence supporting the idea that visual systems do adaptively average information over time consistent with a body of visou-motor work and the general principles of predictive coding (Burge et al., 2008 Harris & Wolpert, 1998; Knill & Pouget, 2004; Friston, 2010). Here we turn our attention to questions relating to the nature and use of past information used in predictive coding more directly.

Predictive coding as a dynamic process.

An interesting feature of predictive coding is its portrayal of the visual system as a dynamic engine of optimal predictions (Clark, 2013; Friston, 2010). In predictive coding, the visual system contains a series of hierarchical models that contain representations of the statistical regularities of the external world (Jun & Chong, 2016). To keep internal representations optimal, sensory systems actively extract information from the world over multiple timescales and levels of complexity constantly trying to maximise extraction of task relevant information from the world (Aitchison & Lengyel, 2017; Clark, 2013; de-Wit, Machilsen, & Putzeys, 2010). However, while there is little doubt that predictive coding considers the visual system capable of making predictions based on a number of different levels of temporal regularities there are a number of outstanding questions and issues related to such ideas. One important issue relates to establishing the ability of the visual system to use complex conditional and sequential relationships that evolve over time to make predictions**.**

The use of sequentially based conditional relationships in making predictions in daily life.

Sequential and conditional relationships occurring within the world are potentially a rich source of information that if applied would offer substantial benefits when making predictions under uncertainty (de Lange et al., 2018). Examples of such relationships and how they might be applied are very common. If you observe a bird flying across the sky at dusk the visibility of the bird is variable

due to decreased light and the way birds behaviour exhibits stochastic elements. Making a prediction about where the bird might be in the next few seconds, might therefore be problematic, as potentially there are a number of possible outcomes. Fortunately, events in the world do not generally occur completely randomly meaning we can apply previous knowledge to solving the problem.

One way to predict the future position of the bird is by using our previous experience that the position of flying birds do not suddenly change from one area of the sky to another. They generally follow a trajectory in which its transition from time point n to n+1 is conditional on its previous positions at time points n-1, n-2, n-3, n-4… as it flies meaning that we can use this transitional information to predict where it will be at a future position. The same principle applies to predicting how events might occur over time on our way into work. If you walk through a busy subway station you need to negotiate walking through lots of people all going in different directions heading to different exits and entrances to subway lines to get onto your required train. Here, there are potentially many different paths people could take making predicting where people might be heading uncertain. One way to predict the position of other people here though is by combining sequential information about how people will transition from the current to a future position based on the transitions between previous n-1, n-2, n-3, n-4… time points combined with knowledge of the exits and entrances of the station. By using this information we can potentially more reliably predict where people are heading. However, while it appears that we have the ability to make predictions about events based on sequential and conditional relationships it is also possible that people do not use such relationships and predict based on more simple relationships between events and stimuli the world.

The use of associative or cued relationships in making predictions in daily life.

Another type of information that can be used to predict the behaviour of stimuli can be considered to be simple associative or cued relationships between events and stimuli in our surroundings (Chun, 2000; Knill, 2007; Kok & Turk-Browne, 2018). For example, if we are sitting watching television and the doorbell rings we can make a prediction that someone is at the door. Importantly, this type of prediction, while still a good way to make a prediction does not require the integration of sequential information and is more of an associative relationship based simply on the cue provided by the doorbell. Similar types of associative relationships exist in many other instances such as on our commute into work. If you are traveling by train into work and the train begins to slow we can take this as a cue that the train is coming into a station and if it is the station we need to get off at can prepare to alight thus saving us time in getting off the train. Again, though the prediction that the train is coming into a station is in this instance is based on the cue provided by the slowing train and while still a good way to make a prediction does not require any knowledge of sequential information of how events unfold over time.

Using the frequency of events to make predictions.

A further type of temporal regularity that can be applied when making a prediction is the frequency of events (de Lange et al., 2018). If we again think of our commute into work we can picture how we can use frequencies of events to make a prediction. If we wish to alight a train we need to know which side the door will be relative to the platform. In many situations, the door will be on the same side on each stop. Say for instance, that the door facing the platform is on the left on 80% on stops then we might use the frequency of times the exit is on the left as a basis for predicting the side of the door on the next stop. Also, if we think about how we perform simple day to day tasks such as opening a door which we have done many time we know that in the vast majority of occasions a door handle opens downwards. This means that the prediction will be that the door handle will indeed open the door if we press the door handle down. However, while the frequency of past events to make a prediction can be a good way to make a prediction as with the use of cues it does not require any use of conditional relationships or sequential information.

Do we use complex sequential conditional relationships to make predictions or do we simply use associative cues or frequencies of events to make predictions.

We have outlined three sources of predictive information that can potentially be used by the visual system to make predictions. However, in many cases current predictive coding experimental designs can be considered somewhat limited in testing whether people can use sequential and conditional information to make a prediction. In many cases, predictive coding experimental designs test more simple cue based predictions or present a stimulus more often in a way that that can be thought of as biasing a particular outcome. Furthermore, there is support from a number of studies from outside predictive coding literature that has shown when presented with sequences of stimuli that contain conditional relationships, participants do not apply them behaviourally when making a decision which clouds the understanding of exactly how predictive information is used in perceptual decisions.

Examples of limited experimental design in predictive coding literature.

One common experimental paradigm used in predictive coding neuroimaging studies that relies on cues to induce predictions is illustrated by Kok, Jehee & de Lange (2012). In this study, the aim was to test how predictions about an upcoming stimuli modulated neural activity and the internal sensory representation of expected and non expected stimulus. Stimuli consisted of gratings presented at two approximate orientations ~45° and ~135° in orientation blocks and at two contrast levels (high and low) in contrast blocks of which types were shown in pairs separated by a blank screen for 500 ms. To induce expectations, the first of the stimulus pair was preceded by ascending or descending auditory tone, which cued the orientation or contrast level of the grating stimulus or the contrast on trials with 75% accuracy. On presentation of the second grating stimulus, participants judged whether the current grating was clock wise or anti clockwise of the first grating stimulus or whether the previous grating

stimulus had a higher or lower contrast than the previous. Analysis focussed on neural activity on the first trial stimulus pair during validly cued trials and none validly cued trials. Consistent with some ideas of predictive coding, imaging analysis showed that neural activity as a whole was reduced (repetition suppression, (Summerfield, et al 2008) during validly cued trials but multivariate pattern analysis also revealed that the internal stimulus representation itself was actually enhanced. This result was interpreted as higher level expectations decreasing neural activity as a whole but also acting to improve perceptions by sharpening or boosting the bottom up stimulus representation in early visual cortex. While this was an interesting result and theoretical interpretation there are a number of issues with the use of cues as in Kok, Jehee & de Lange (2012). Experimental designs, such as in Kok, Jehee & de Lange (2012) while ideal for their purpose, do not test the constant changing and updating of predictions based on sequential information as they might occur in the environment. Furthermore, the use of cues can be also considered to set up more of an associative relationship as opposed to a predictive relationship. One could think of this as hearing a doorbell ring. When a doorbell rings we do not necessarily need to make a prediction about an event and we may just associate the sound of the doorbell with someone being at the door and respond accordingly. The type of cue/prediction paradigm in Kok, Jehee & de Lange (2012) is a common type of experiment also used in Kok & Turk Browne (2018).

Kok & Turk Browne (2018), examined the role of the hippocampus in making cross modal (auditory and visual) predictions and its potential role with providing predictive information to visual cortex. Stimuli consisted of two shape pairs separated by a blank screen for 500 ms. To induce expectations, the first of the stimulus shape pair was preceded by ascending or descending auditory tone which cued a particular shape with 75% accuracy over trials. Analysis focussed on neural activity and stimulus representation in the hippocampus and visual cortex on the first trial pair during validly cued trial and none validly cued trials. Results found that the stimulus representation in visual cortex was dominated by the current shape regardless of valid or invalid cues but in the hippocampus shape representation reflected the validly cued shape. The cued shape representation in the hippocampus should be said is a very interesting result indicating support for the predictive coding notion of a hierarchical system of learnt regularities with the hippocampus at the top of the hierarchy. However, again this type of paradigm, could be considered to set up associative relationship between tone and shape as opposed to truly predictive relationship. Furthermore, by making one stimulus much more common by presenting it on 75% of trials it leaves open the repetition priming interpretation in which responses are modulated by the frequency of a presentation as opposed to any learnt sequential or conditional information.

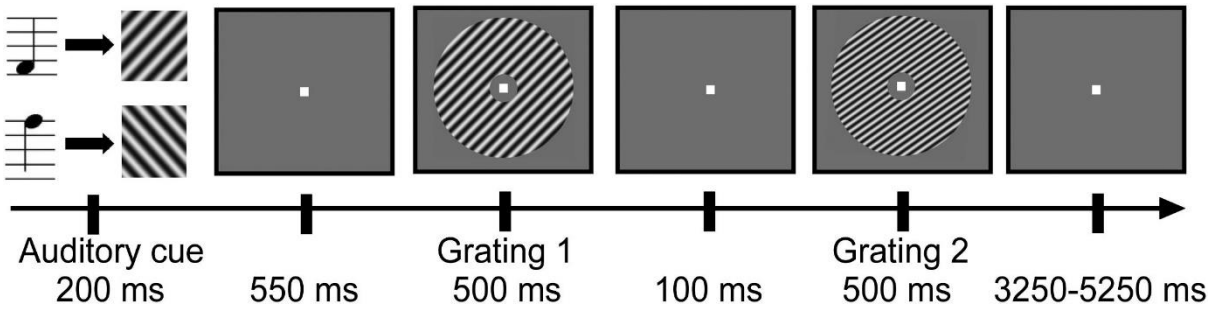


Figure 1. Kok, Jehee & de Lange (2012) stimuli. On each trial participants were presented with two grating stimuli presented consecutively. The second grating differed in termed of either orientation or contrast. In separate blocks, subjects performed either an orientation task ("Was the second grating rotated clockwise or anticlockwise with respect to the first?") or a contrast task ("Was the second grating of higher or lower contrast than the first?"). The grating stimuli were preceded by an auditory cue, which predicted (with 75% validity) the overall orientation of the gratings (~45° or ~135°) (taken from Kok & Jehee, 2012 permission from Dr Peter Kok)

Questions about whether people apply sequential information in perceptual decision making

A further problem for predictive coding's ideas about the use of conditional relationships in sequences of stimuli, is that in some cases evidence that people apply them in perceptual decision making in a way that is contradictory with predictive coding. At the heart of predictive coding is the idea that internal models contain a representation of the conditional relationships between stimuli and events in the world observed over time to predict future events (Clark, 2013; Friston, 2010; Jun & Chong, 2016). The use of the most likely stimulus is proposed to reduce prediction error by representing and applying what the brain has learnt about the environment to what will happen next which should have distinct benefits in terms of reaction times. This is because, according to predictive coding when predictions match sensory signals no further processing is required as the previous prediction remains valid (Rao & Ballard, 1999; Summerfield & Koechlin, 2008). Under the principles of psychophysical theory, when the brain is using less resources reaction times are decreased (Henry, 1980; Klapp, 2010; Maslovat, Klapp, Jagacinski, & Franks, 2014). Alternatively, when predictions do not match sensory inputs a prediction error is produced using more resources making reaction times slower. Indeed, this relationship is central to predictive coding and has led some to state that prediction error and reaction time are intimately related (Summerfield & Koechlin, 2008). However, there are nonetheless studies that have reported that this relationship is not so clear cut in terms of actual predictions in the forwards sense.

Maljkovic & Nakayama (1994) classic priming study, assessed the factors influencing pop out in visual search tasks. In the paper, they compare number of top down and bottom up factors that could influence pop out in a simple visual judgment tasks. Crucially, in terms of predictive coding these factors included the conditional probability of the colour of presented stimuli against priming or a simple short term visual memory trace tuned to consecutive or more frequency of trials. The task of

participants was to detect as fast as possible the odd coloured diamond shape (see figure 2 below). Three participants carried out a number of blocks in which the transition probabilities of whether the colour of the odd coloured shape would repeat or switch was manipulated. These probabilities were 0, 0.1 0.5, 0.7, 0.9 and 1. Crucially, these probabilities set up two totally predictable sequences and one totally unpredictable sequence.

The two predictable sequences were found in 0 probability blocks that never switched and probability of 1 blocks that always switched while the unpredictable sequence was found in blocks that that had a 0.5 probability which switched at chance levels. This experiment was subject to two competing theoretical predictions. One theory was termed 'stimulus expectancy' in which observers perform better in proportion to the level of certainty in the sequence and a 'short term memory' or priming hypothesis in which the number of consecutive trial determines performance as the memory trace strengthens over repeated trials. If the stimulus expectancy hypothesis was correct, then reaction times should have been faster and roughly equal for 0 and 1 probability stay/switch blocks as both of these blocks change in a totally predictable way thus have equal levels of uncertainty (zero) with reaction times slowest for the 0.5 probability blocks which had the highest levels of uncertainty as colours changed at chance levels. On the other hand, if the priming hypothesis was correct then predictability of stay/switch would still have an effect but it would only lead to faster reaction times in the 0 probability stay/switch blocks. This is because as colours in these blocks always repeated, as opposed to always switching as in the probability of one blocks, in which trials always switched.

Results supported the priming hypothesis. In 0 probability stay/switch blocks reaction times were fastest and slowest for one probability stay/switch blocks with 0.5 in the middle. This result supports the idea that the brain can use sequential information but is inconsistent with predictive coding. This is because while it appears that the conditional probabilities did play a role in responses they only improved responses when the conditional probability predicted events staying the same and not for when things changed. In predictive coding, the emphasis is on predicting the next event with reaction times taken as a measure of prediction error. However, this result supports the idea that sequences were not reducing prediction error on the next trial and asks questions about predictive coding's relationships with behaviour.

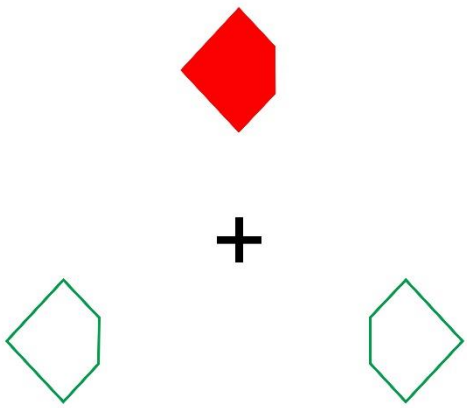


Figure 2. Maljkovic & Nakayama (1994) stimuli. The task of the participant was to judge whether the odd coloured diamond was cropped to the left of right with the colour of the stimulus the manipulated quantity with the other two shapes fulfilling the role of distractor shapes

One possible reason relates to the factor the study manipulated. In the study, they manipulate whether the colour of the shape will stay the same or switch but not what the next colour would be. While there were only two colours, meaning that it might have been quite easy to predict what the next colour might have been, this was not what the study tested. It might be that sequential transition probabilities are applied more readily when there is an actual prediction of what the next specific quantity might be as opposed to whether something will stay the same of switch. Another explanatory factor might be due to the way predictive information interacts with decision making. While, it is supported that predictive information does influences perception (Bergen & Jehee, 2019; Chopin & Mamassian, 2012; den Ouden, Kok, & de Lange, 2012; Jun & Chong, 2016) and neural activity (de Lange et al., 2018; Kok & Turk-Browne, 2018; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008) it is unclear how prior knowledge effects the decision process. It may be that the brain learns the sequential probabilities of events but does not apply them in an optimal way all the time (Feher da Silva, Victorino, Caticha, & Baldo, 2017; Newell, et al 2013; Unturbe & Corominas, 2007). Furthermore, is may also be possible that non rational cognitive biases influences the decision making the outcome non optimal (Kahneman, Slovic, & Tversky, 1982). One study that perhaps inserted a sequence into their stimulus presentation that favoured prediction of what would happen next in perhaps a clearer way than the stay/switch paradigm used in Maljkovic & Nakayama (1994 is Jones & Pashler (2007).

Jones & Pashler (2007), set out to test the idea that the mind is tuned for detecting forward relationships. In regard to predictive coding this can be considered theoretically important. Predictive coding very much paints the brain as a predictive machine constantly trying to anticipate what will

happen next (Clark, 2013; Friston, 2010, 2018b) and if this is not the case then would invite questions about both predictive coding and more general ideas about the idea of predictive brain. To test this theory, Jones & Pashler (2007) inserted a specific type of conditional stimulus sequence that favoured the detection of forward stimulus relationships called a Markov chain (see figure 3). A Markov chain (see Rønn-Nielsen & Hansen, 2014 for an explanation of Markov chain theory) is a stochastic process in which the conditional probability of a switch from the current value, often termed state, to another state is dependent upon only the present state. In the experiment, Jones & Pashler (2007) first trained participants on 600 images to participants comprised of 8 shapes (see figure 3). Participants were asked to attend to the shapes but not told the purpose of the study or what task would follow.

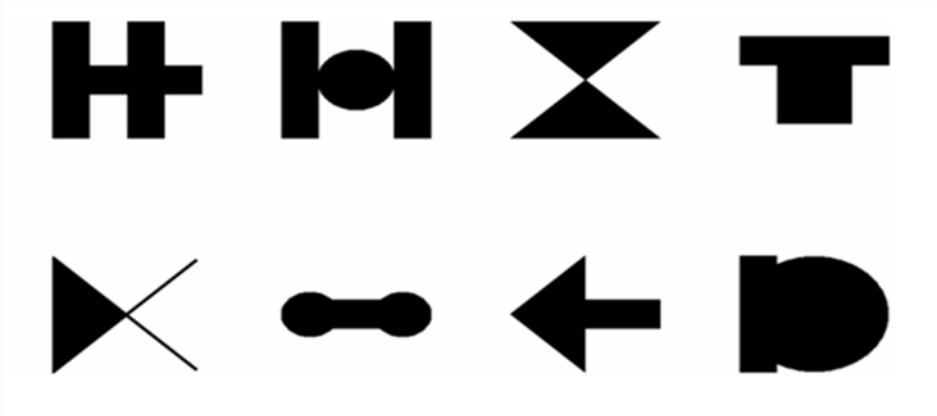


Figure 3. Jones & Pashler's (2007) 8 shapes (Top left is shape 1, bottom right is shape 8)

Following this training phase, subjects were informed they would be tested on the order the shapes were presented and asked two questions-a 'prediction question' in which they were asked which shape was most often presented immediately after a specific shape and a retrodiction question which was which shape was most often presented immediately after a given shape. To clarify, in predictive questions participants were asked what the next shape after the current shape would be while retrodictive questions asked they were asked what the shape before the current shape was. The study hypothesized, that if prediction is a fundamental principle of brain function, then we might expect prediction questions to see better performance as forward temporal relationships are preferentially detected. Alternately, if backward temporal relationships are preferentially detected then they would expect to see greater accuracy in retrodictive questions. Results reported some interesting findings. Overall, accuracy (36%) was well above chance (12.5%) meaning participants appeared to be able to learn the stimulus sequence. However, there was no significant advantage for prediction over retrodiction. Accuracy was 39% for prediction, and 33% for retrodiction. Given that participants were not told anything about the nature of the task during training Jones & Pashler (2007) conceded that

there was no real task relevance for detection of forward temporal relationships during the training phase as participant had not been informed about the nature of the task and may not occur in purely incidental learning tasks. To test this possibility they conducted a second experiment in which subjects were told to expect either the retrodiction or prediction question. However, again results reported no improvement in accuracy and in summary results can be considered at odds with predictive brain theories.

There are a number of possible explanations for the result indicating that forward temporal relationships were not favoured. One could be that manifestation of favouring forward directed relationships is a more implicit process and lower order phenomena than used during explicit judgments as in Jones & Pashler (2007). Evidence suggests that the expression of favouring forward relationships becomes more prevalent when people repeat sequences of motor responses as shown in studies of implicit sequence learning (Nissen & Bullemer, 1987). This might be considered logical as motor actions are always forward in time. In this study, the task required an explicit reporting of the next and previous stimulus. Another potential explanatory factor is that in tasks that require explicit decisions, people often exhibit differing uses of probabilities in sequences of stimulus that are at odds with what might considered optimal. For example, one type of suboptimal use of conditional probability is where people even if they know the probabilities believe that one event is more likely to occur because it has not occurred for a time often termed the gamblers fallacy (Urteaga & Wiggins, 2018). For example, in roulette if the ball has not landed on red for 5 spins people think it is more likely to happen even though the odds of red are always 50%. Another type of non optimal use of sequential information is probability matching. Probability matching refers to the phenomena that when presented with a stimuli that changes with a higher probability, say 0.7 to one state (A) than another state (B) at 0.3 and asked to judge what the next state should be instead of always choosing the most probable state, responses match the overall probabilities of the sequence meaning they respond with state A, 0.7 of the time and state B, 0.3 of the time (Gaissmaier & Schooler, 2008; Koehler & James, 2010; Newell et al., 2013). Finally, aside from non optimal use of conditional information it is perhaps possible that sequences of stimuli work to provide a more simple associative relationship, as opposed to a truly predictive relationships in which the direction of events is less important than the link between stimuli as seen in association learning studies (Rudin, et al, 2008; Schultz, 1998; Wolfensteller & Ruge, 2011)

State transitioning from (current trial)	A	B
A	0.2	0.8
B	0.7	0.3

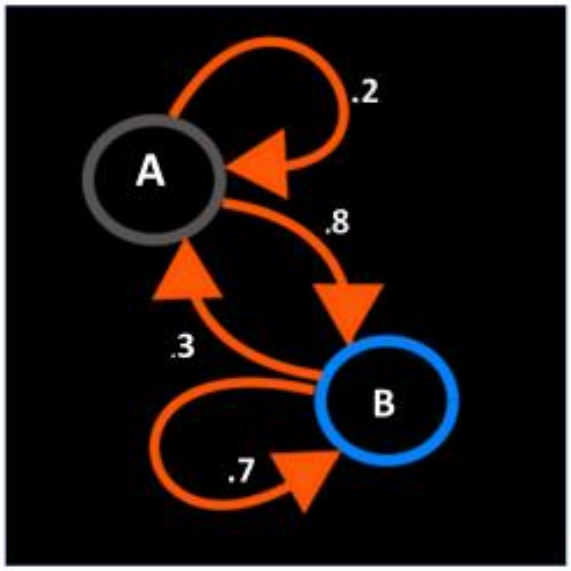


Figure 4. Markov chain and associated transition matrix- Markov chains, named after the Russian mathematician Andrey Markov, are a stochastic process that operate by sequentially transitioning from one state (some value or situation) to the next. For example, if you made a Markov chain about the weather you could include, rain, sun, fog and cloud which would form a 'state space' or a list of all possible states . A markov chain would tell you the probability of transitioning from one state to any other state in the state space e.g. given that the current state is rain the chances of transitioning to sun, cloud, fog or remaining in rain. Above we have a Markov chain with two states, A and B where there are four possible transitions A-A, A-B, B-A and B-B. Next to the illustration we have a table which shows how Markov transitions are normally detailed termed a transition matrix. Both show that if we are in state A the probability of transitioning to state B is 0.8 and 0.2 for staying in A, if the current state is state B the probability of transitioning to state A is 0.3 with a 0.7 chance of remaining is state B. Transition probabilities between states must sum to one. Importantly although the chances of transitioning to any state within a state space are totally conditional on the current state they represent the outcome of a sequence of events that have provided these probabilities. Here, say the Markov chain changed state every two seconds and we observed transitions for an hour. We could then think of the Markov chain as a model of these transitions in which all that needs to be known to predict the next transition is the value of the current state. This aspect of Markov chains provides a 'memoryless' quality as the state of the system at a future time is decided entirely by the state at the current time only. This efficiency has made Markov chains an especially attractive model in a number of computational predictive coding models and empirically relates Markov chains to the Kalman filter which also makes predictions based only on the current state assuming an underlying Markov process in continuous time.

5.2.1 The current chapter: theoretical motivations, aims and hypotheses.

The current chapter aims to assess a number of questions relating to predictive coding's portrayal of the brain as a dynamic system able to produce predictions based on conditional relationships in sequences of stimuli. Previous predictive coding work has not fully tested this idea and often used simple cue type association paradigms or provided results that can be interpreted as being more influenced by frequencies of presentation. Furthermore, other studies that have inserted sequential probabilistic relationships found that either they did not improve behaviour or apply them in a strictly forward manner. However, the outlined studies we discussed to illustrate some of these points from Maljkovic & Nakayama (1994) and Jones & Pashler (2007) while not designed for testing these aspects of predictive coding (or predictive coding at all) still provide some useful ideas that can be used for testing the ability of visual systems to make predictions based on conditional relationships in sequences of stimuli. Here we incorporate ideas from Maljkovic & Nakayama (1994) and Jones & Pashler (2007) and extend their designs. We introduce an experimental design that like Maljkovic & Nakayama (1994) manipulates the conditional probabilities of stimulus sequences but inserts a more forwards looking relationship by inserting a Markov chain into our stimulus sequence as in Jones & Pashler (2007), also applying a reversal of the transition probabilities. To assess whether information contained in sequences of information favour forward looking relationships we inserted omission trials into our sequence of stimuli and ask participants what the stimulus should be. We measure reaction times and also record the percentages of presses made on omission trials relative to previous trials to see how people use sequential transition probabilities in these instances.

To provide theoretically driven analysis, we make experimental hypotheses based on predictive coding theory and compare these ideas against those of short term memory or priming/repetition effects. In terms of reactions time if people do use the sequential transition probabilities in our stimulus sequence in a way consistent with predictive coding then we would expect faster reaction times for more predictable stimuli. This is because according to predictive coding theory, when predictions are valid to the internally modelled sequence probabilities there should be less prediction error leading to a faster response (Summerfield & Koechlin, 2008). Alternatively, if the number of repetitions is the more important factor, as in Maljkovic & Nakayama (1994), as we held the total number of screen positions constant, there should be no preference for predictable transitions and no difference in reaction times. In terms of responses on omission trials, we can also make two competing predictions. In predictive coding, predictions make optimal use of transition probabilities. This means that the response should always be to select the most probable stimulus. If this is the case, we would expect that, on omission trials, participants will always select the most probable transition given the previous stimuli. On the other hand, if repetition is the more important factor as in priming then no preference should be given to the most probable stimulus.

5.3 Methods.

Ethics

All calibration and experimental procedures were approved by the University of St Andrews Teaching and Research Ethics Committee. All participants gave informed consent.

Participants.

A total of 10 participants took part in our main experiment (6 female) mean age 33 (range 19-42). All participants were recruited as volunteers from St Andrews SONA recruitment database. Participants had not taken part in any previous experiments detailed in the current thesis.

Stimuli design software and presentation details

In all experimental and calibration procedures visual stimuli were created in MATLAB (The Mathworks Inc, Natick, MA) and presented using PsychToolbox (Brainard, 1997).

Main experiment-inserting and testing the use of conditional transition probabilities

Our main experiment aimed to manipulate the conditional probabilities of the way our stimulus changed position over trials in a more dynamic and sequential way than previous predictive coding studies. The aim of this manipulation was twofold with each aim forming the basis of specific analyses. The first aim was to analyse whether participants were able to learn the Markov transition probabilities and to test whether applying such information would improve behaviour in the shape of decreased reaction times. The second aim was to assess whether participants would use the transition probabilities in a forward direction by analysing the presses made by the participant on omission trials relative to the previous presented state.

Experimental stimuli

Our experimental stimuli were Gabor patches. In most trials Gabors would appear in one of three screen positions over trials; centre centrally and slightly towards the top of the screen (X=0 °, Y=5°) right slightly above centre (X= -3.53 °, Y=3.53°) and left and slightly above centre (X= 3.53 °, Y=3.53°). Gabor stimuli had a radius of 8 visual degrees and had a spatial frequency 0.5 cycles per visual degree and were presented at 25% contrast (Michelson (see figure 5 below) but on some 'omission' trials no stimuli would be presented and participants instead viewed a blank screen.

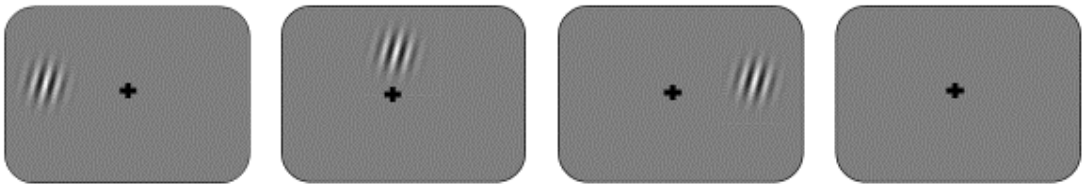


Figure 5. Basic details of experimental stimuli. Trials either presented Gabor stimuli in one of three screen positions or did not present a stimulus at all.

Manipulation of transition probabilities-Markov chains

Our experimental blocks contained two Markov chains sequences. One Markov chain we term 'forward matrix' (see table 1) and one we call 'backwards matrix' which was the exact mirror of the forward matrix (see table 2). Blocks consisted of 100 trials. Importantly in all blocks, during the first 80 trials Gabors changed between screen positions and omissions transitioned between trials in a way determined by the forward matrix transition probabilities with switches between screen positions now defined as states A, B, C and D. State A relates to the left sided Gabor, state B the centrally positioned Gabor, state C the right sided Gabor and state D which were omission trials. In the other 20 trials within a block, Gabors changed between screen position based on the backward matrix transition probabilities to states we term E, F, G and H (state E relates to the left sided Gabor, state F, the centrally positioned Gabor, state G the right sided Gabor and state H which was a blank screen) (see figure 7 below). Importantly, within blocks transition probabilities between states were structured in such a way that all screen positions were presented an equal number of times (see table 3 below). This was to make each screen position equally likely which was crucial to testing the priming explanation of reaction time. Here we emphasize that on both forward and reverse matrix phases within blocks the same screen positions were always used. To clarify this means that states A and E both correspond to Gabors presented at the left of the screen, states B and F to Gabors presented centrally, states C and G to Gabors presented at the left of the screen and states D and H presented blank screens (see figure 6 below). The reason why we give the same screen positions different 'states' is because we need to differentiate trial states for analysis purposes.

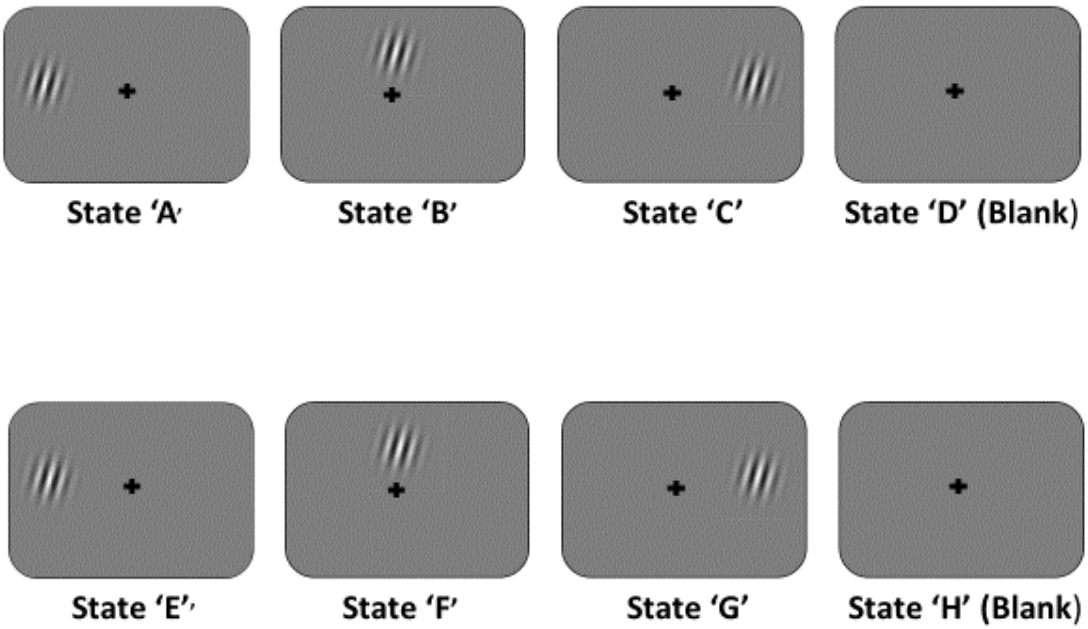


Figure 6. States relative to screen position in the forward and backward matrix phases of blocks. In the forward matrix phase (top) we assigned the following screen positions to the following states. State A left and slightly above centre, state B slightly above the centre of the screen, state C to the right and slightly above centre and D no stimuli presented. In the backward matrix phase of clock (bottom) we assigned the following screen positions to the following states. State E left and slightly above centre, state F slightly above the centre of the screen and state G to the right and slightly above centre and state H no stimuli presented.

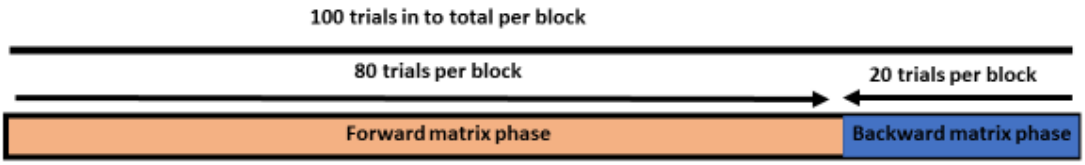


Figure 7. Trial structure in each block-here we illustrates the trial structure of our block. For 80 trials, states switched in a way determined by our forwards matrix and then on trial 80 trials switched in a way determined by our backward matrix**.**

State Transition matrices.

The transition probabilities between states were determined by our forward and backward matrices which are what are known as Markov chains. Markov chains are defined as a stochastic process in which the transition to the next state is determined entirely on or is conditioned by only the current state. Within Markov chains, each state is associated with a set of probabilities for switching to any of the other possible states. By building in and altering transition probabilities which are which are presented below in tables 1 and 2, we made certain transitions between states more probable than others over trials and therefore define these transitions as predictable.

Table 1 State transition probabilities 'forward' matrix

	State transitioning to (next trial)			
State transitioning from (current trial)	A	B	C	D
A	1/9	6/9*	1/9	1/9
B	1/9	1/9	6/9*	1/9
C	6/9*	1/9	1/9	1/9
D	1/3	1/3	1/3	0

*denotes predictable transition pairing

Our forward and reverse matrix tables (table 1 above) show exactly where the predictable state pairs exist (denoted by a star). In forwards matrices, if the current state is state A the predictable transition on the next trial is to state B which occurs 6/9's of the time as opposed to remaining in state A or changing to states C or D (omissions) which all occur with equal probability (1/9). If the current state is B, then the predictable transition on the next trial is to state C, which again occurs 6/9's of the time with an equal probability of 1/9 to states A, staying in B or switching to D. Lastly when the current state is D (blank) there is a zero probability of staying in state D (so blanks are never repeated) and a 1/3 chance of transitioning to any other state. In the backward matrix phase of blocks the state transition probabilities are the reverse from the forward matrix as shown below in table 2.

Table 2 State transition probabilities 'backward' matrix

	State transitioning to			
State transitioning from	E	F	G	H
E	1/9	1/9	6/9*	1/9
F	6/9*	1/9	1/9	1/9
G	1/9	6/9*	1/9	1/9
H	1/3	1/3	1/3	0

*denotes predictable transition pairing

In the backward matrix phase of blocks if the current state is state E the predictable transition on the next trial is to state G which occurs 6/9's of the time as opposed to remaining in state E or changing to states F or H which all occur with equal probability (1/9). If the current state is F then the predictable transition on the next trial is to state E which again occurs 6/9's of the time with an equal probability of 1/9 states, staying in F or switching to G or H. When the current state is G the predictable transition is to state F which happens on 6/9's of the time and a 1/9 chance of switching to states E, F or H. Finally, if the current state is H (blank) there is a zero probability of staying in state H and a 1/3 chance of transitioning to any other state. Crucially, although some state transitions occurred more often than others, transition probabilities were designed in such a way that they controlled for frequency of presentations. In all blocks, the frequency of screen positions presented over the 100 trials was always the same. This meant that stimuli were presented approximately 30 times in each screen location with 10 blanks or omission trials also always being presented (see table 3 blow).

Table 3 approximate number of screen positions presented per block

Stimulus presented	Total number of presentations
Right	30
Centre	30
Left	30
Blank	10

Procedure.

Participants were seated 57cm from a computer monitor. Participants were told that they would be presented with a round grating stimuli that would appear in three screen locations, left, middle and right and that during trials they should try to focus on the fixation cross in the middle of the screen. We asked participants to press a button on a button box that pertained to the one of each of the screen positions as fast as possible when they saw the stimulus i.e. press the left button if the stimulus was on the left and the centre button if the stimulus was in the centre. Importantly, we also informed them that although most of the time stimuli would be clearly visible, on occasion this would not be the case and on some trials the stimulus would be very difficult to see and during these trials they should press the button they think applies to where the stimulus should be as fast as possible. For clarity, on these trials stimuli were not presented, forming our omission trials. Each stimulus that was presented appeared for 200ms seconds with the gap between trials drawn from an exponential distribution of between 100-500ms to prevent participants from learning the time gap between trials which could potentially have confounded our reaction time data. In the forward matrix phase of blocks 80 trials were presented. On the 80^{th} trial the Markov chain was then reversed and trials followed the backward matrix transition probabilities. Initial states were randomly selected. Prior to the start of experiments participants practiced the task until they felt comfortable and the researcher felt they understood the task.

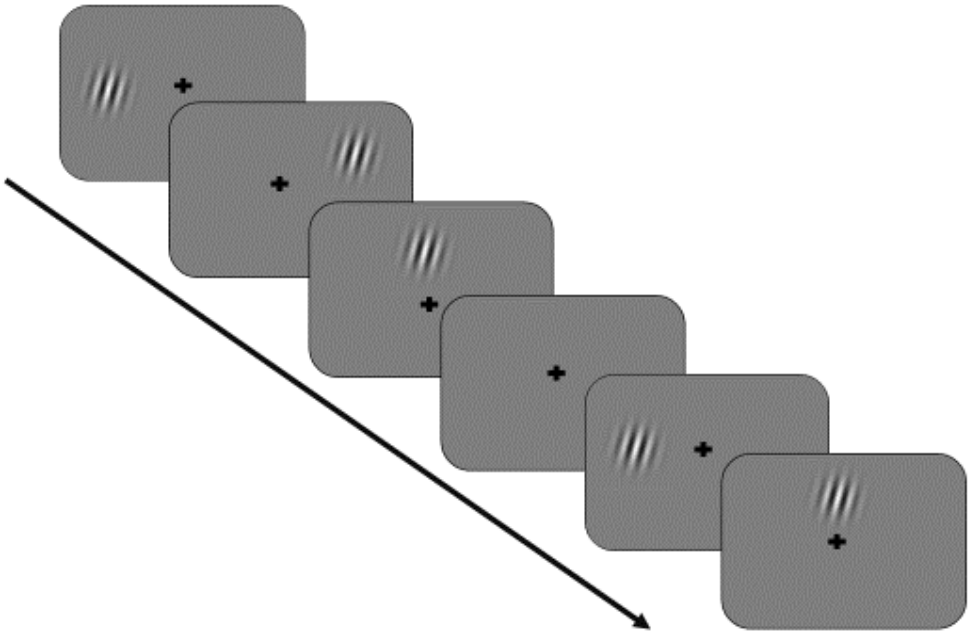


Figure 8. Stimulus design and event sequence. On each trial either a stimulus at one of three orientations or a blank screen termed 'omission trial' was presented for 500ms.Stimuli were presented immediately after one another with a small inter trial interval drawn from an exponential distribution of between 0.1 and 0.5 seconds. The task of participants was to signal a judgment as quickly as possible about the perceived position of the Gabor stimulus. To allow for a response on omission trials we told participants that on some occasions stimuli would be very difficult to see and they should simply press where they think the stimuli should be.

Trial numbers and blocking.

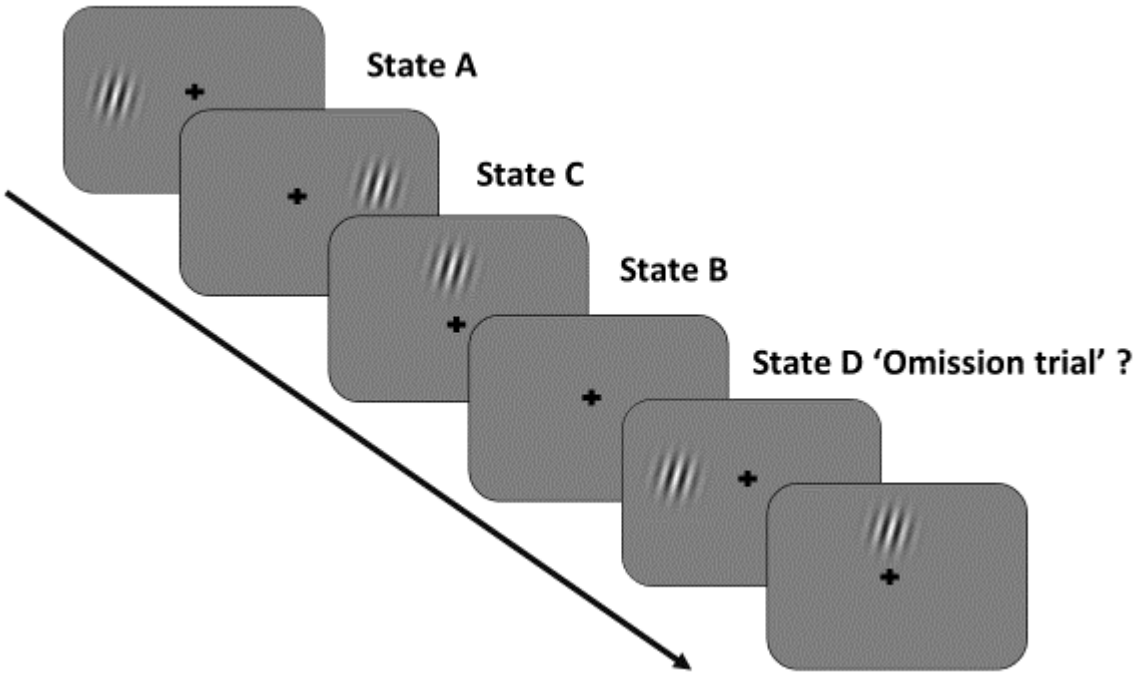
Each block contained 100 trials. Within each block 80 trials were in the forward matrix phase and 20 in the backward matrix phase. Each subject completed 9 blocks over two one hour sessions. This gave 2700 trials in total. Blocks always began with forward matrix phases

Reaction time analysis: influence of transition probabilities

We analysed reaction times for all visible state transition pairings (all non omission trials) in both forward and backward matrix phases of blocks. Here we isolated three key independent variables that could have potentially have an effect on reaction time which represented the dependent variable. These independent variables were 'predictability' as defined by probability of transition between states, stimulus screen position/state and forward and reverse transitions phases (except for transitions to omission trials which form the basis of analysis two). We analysed the effects of these independent variables on reaction times in the following way. Analysis of the effects of our dependent variables on reaction times is performed using a three way repeated measures ANOVA to test for both main effects of predictability on reaction times and interaction effects between predictability and both current state and forward and reverse transitions phases. For post hoc analysis of any significant differences in reaction times between transition pairings we use paired samples t tests (Bonferroni corrected) and report 95% confidence intervals.

Omission trial analysis-influence of previous trial on omission trial response.

Here we analysed if omission trials were influenced by our transition probabilities. This is performed by analysing responses on an omission trials in relation to the previous trial state. For example, during forwards matrix phases of blocks if an omission trial was preceded by state A would participants more often press to signal based on most probable transition from A which would have been B more often than other states. Alternatively, in the reverse matrix phase if state E preceded an omission trial would participants signal the most probable which would have been state G. The same principle was applied for all omission trials. Here our dependent variable is the percentage of times a screen position was selected with the independent variable the preceding state. We analysed percentages pressed on omission trials relative to the previously presented state using a repeated measures ANOVA. For post hoc tests on any significant results we use paired samples t tests (Bonferroni corrected) and report 95% confidence intervals. Here we do also report reaction times but this is for general comparison with reaction times with non omission trials.



	State transitioning to (next trial)			
State transitioning from (current trial)	A	B	C	D
B	1/9	1/9	6/9*	1/9

Figure 9. Clarification of omission trial analysis-Here we provide a hypothetical trial sequence during a forward matrix phase of a block. In this sequence the states presented begin with state A, then state C and then state B. On the fourth trial an omission trial (state D) is presented. If participants use the transition probabilities to make a judgment on the omission trial then because the previous trial was state B then they should signal the current state as being the most probable position which would be state C (6/9) and signal this by pressing the button that corresponds with that state. Over trials this would be represented in the overall percentages for omission trials preceded by state B which would be higher than the percentages for the other possible states. However, if participants' do not integrate the transition probabilities when judging on omission trials then percentages for all possible states would be the approximately equal

5.4 Results.

5.4.1. Influence of predictability, transition matrix direction and screen position on reaction time.

A repeated measures ANOVA tested the effects of predictability, current screen position and the direction of our transition matrices on reaction time. Analysis reported a main effect of predictability between transition pairs on reaction time F (1, 9) = 6.171, p=0.035 and a significant interaction effect of transition matrix direction F (1, 9) = 5.48, p=0.039 (see figure 10 below) No significant effects of current screen position on reaction times were recorded. Mean reaction time for predictable transition pairings were 469.3 ms (SD=639.1, CI [440.7, 498.1]) and non predictable transition pairings 501.7 ms (SD=0.089.1, [475.5, 527.9]). Mean reaction times for forwards matrix transition pairings were 496.3 ms (SD=8331, CI [476.4, 516.1] and for forwards matrix transition pairings 4896 ms (SD= 819.4, CI [470.0, 509.1].

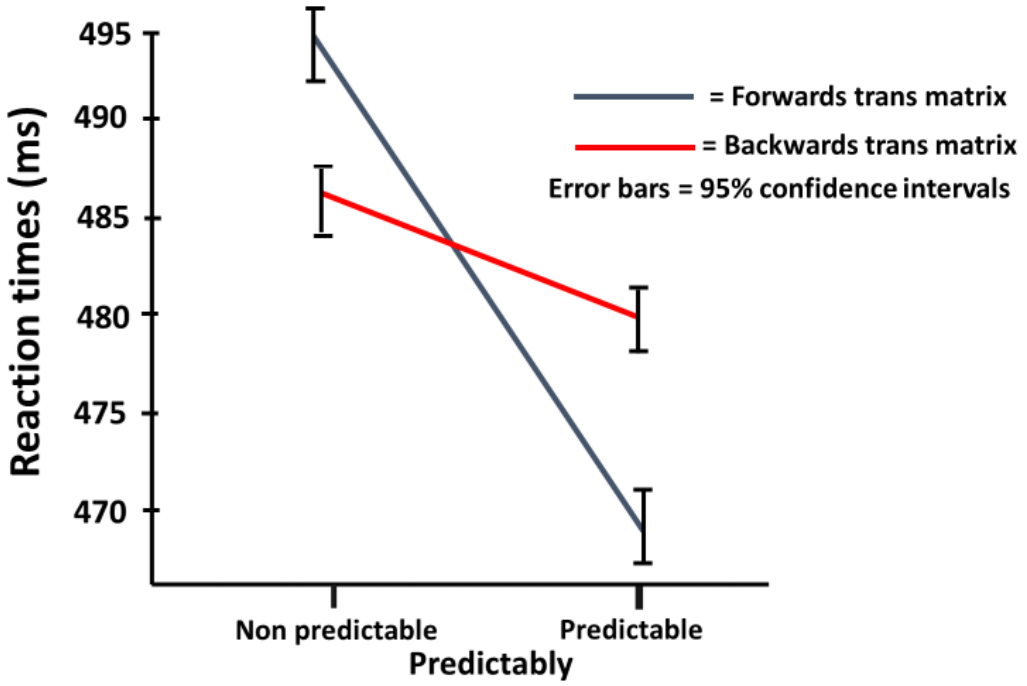


Figure 10. Interaction effects of predictability in transitions pairings and direction of transition matrix on reaction times and 95% CI's. Crossed lines in our interaction effect plot indicates that changes in reaction times depended upon the both the level of predictability and the direction of the transition matrices One of the most interesting aspects of our reaction time data is the significant interaction effect between predictability and the direction of the transition matrix on reaction times. The crossed lines in our interaction effect plot indicates that changes in reaction times depended upon the both the level of predictability and the direction of the transition matrices. This interaction can be interpreted as behaviour having changed due to the switch in the most likely transitions pairings. Conceivably this change in behaviour perhaps indicates dynamic updating of the transition probabilities had occurred even over a small number of trials consistent with predictive coding

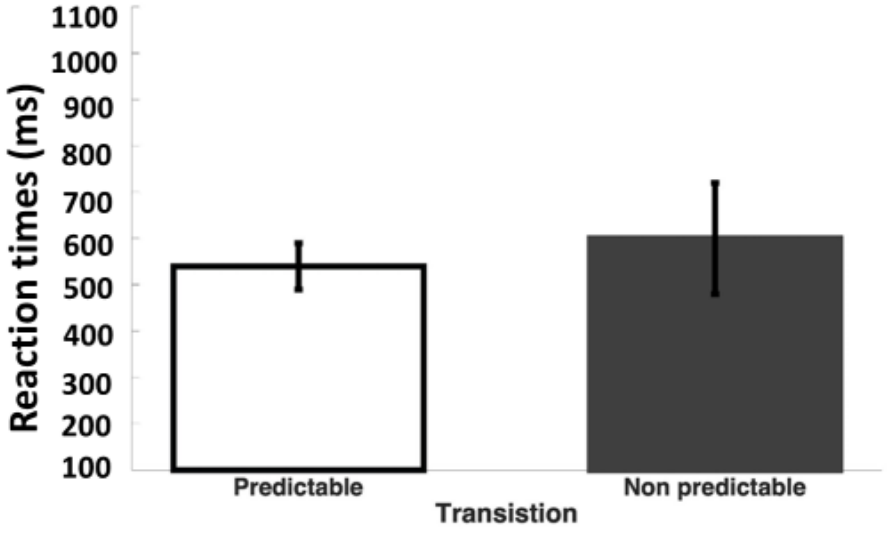


Figure 11. Bar graph showing mean reaction times and 95% CI's for predictable and non predictable transition pairings. Here we observe faster reaction times for predictable transition pairings with a slower and more variable response for non predictable pairings

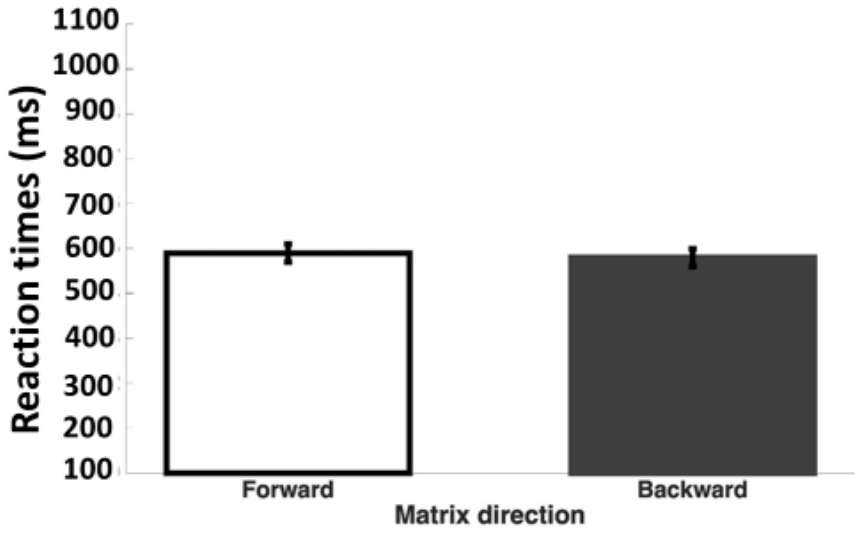


Figure 12. Bar graph showing mean reaction times and 95% CI's for forwards and backward transition pairings. Here we observe that there is no difference between forwards and backwards matrix transition paring indicating that predictability is the most important factor on reaction times.

Post hoc tests on forwards Matrix transition pairings.

We conducted t tests after our ANOVA on reaction times between predictable state transitions (star denotes predictable transition pairings) and non predictable state transition pairings. Results reported there were no significant A-B* (M=0.46, SD=0.06) and A-A (M=0.50, SD=0.10) reported no significant differences in reaction times=0.61 however, paired samples t tests between state transitions A-B* (M=0.46, SD=0.06) and A-C (M=0.50, SD=0.08) did report significant differences with t(9)= -3.513, *p*=0.12. Here we also note that the *p* values for the A-B* and A-A was very close to significance at *p*=0.055. Results for state B transitions were consistent with those from state A transitions. Here we recorded very strong significant differences in reaction times between state transitions B-C* (M=0.46, SD=0.08) and B-B (M=0.50, SD=0.85) with t(9)=-6.813, p<0.00, non-significant results were recorded between B-C* (M=0.46, SD=0.08) and B-A (M=0.48, SD=0.07) but again were close to significance for this pair with a *p* value of 0.059. Results from C state transitions were again consistent with those of A and B with significant differences recorded in reaction times. Paired samples t tests of reaction times between state transitions between C-A*(M=0.47, SD=0.08 and C-C (M= recorded significant differences with t (9) =-2.87, *p*=0.18 again though we only recorded significant differences in one of our comparison as the t test of reaction time differences between C-A and C-B reported non significant differences (see table 6 for means reaction times and 95% confidence intervals for all transition pairings).

Post hoc tests provide added interpretation of our significant ANOVAs results and are consistent with the idea the extraction and use of conditional probabilities can enhance behaviour in a measurable way. Our data also provides support that participants were using the transition probabilities rather than being influenced frequencies of transitions as might be the case with priming effects. If we look at where our significant differences between states lie, although we did not record significant differences for every possible comparison there are significant comparison for all of the possible switches between actual screen positions. For example we recorded significant differences between states A-B. Translating this transition to screen positions this means that participants were faster pressing when the switch was from left to centre. However, we did not record significant differences when the switch was from right to centre (C-B) instead on C transitions the fastest reaction times were from right to left (C-A). Also, on the B transitions (centre to right/left) the fastest reaction times were found when the stimuli transitioned from the centre of the screen to the left of the screen (B-C). This means that we recorded a faster reaction time involving transitions to all possible screen positions and all possible button positions.

Table 4-Mean reaction times(ms) and 95% confidence intervals-forward matrix transition pairings

Transition pairing	Mean reaction time	95% Confidence interval (lower and upper bound)
A-A	500.1	421.2, 571.5
A-B*	465.2	418.1, 515.5
A-C	500.1	446.2, 567.4
B-A	481.2	431.2, 546.1
B-B	502.1	443.1, 569.8
B-C*	468.1	401.2, 521.2
C-A*	471.2	417.7, 532.1
C-B	498.3	448.1, 552.1
C-C	510.9	441.2, 582.1

*Denotes predictable transition paring

In contrast to forward matrix transition pairings, backwards matrix transition pairings reported no effects of predictability between transition pairings on reaction times. While in hindsight this might have been expected due to the smaller number of trials presented during the backwards matrix phases this is not to say that reversing our transition matrix had no effect on behaviour. Indeed the fact the we did not record significant effects despite the fact that the same screen positions were presented and task buttons remained the same indicated that something related to the switch in transition probabilities did modulate participants reaction times. For example, in the forward matrix transition pairings we recorded significant differences between left to right stimulus transitions (A-B) which we

did not observe in the equivalent screen stimulus positions in the backwards matrix transitions (E-F). This change in patterns of reaction times is supported by our significant interaction effect between predictability in transition pairs and the direction of our transition matrices.

Table 5 Mean reaction times (ms) and 95% (CI's)-backwards matrix transition pairings

Transition pairing	Mean reaction time	95% Confidence interval (lower and upper bound
E-E	501.1	442.3, 576.5
E-F	482.4	436.6, 531.2
E-G*	483.4	432.3, 534.4
F-E*	472.3	413.6, 546.3
F-F	496.7	432.6, 552.3
F-G	491.2	421.1, 571.2
G-E	489.8	421.2, 535.5
G-F*	492.3	446.3, 538.3
G-G	513.4	442.3, 582.1

*Denotes predictable transition paring

5.4.2 Omission trials: influence of previous trial on omission trial responses.

A repeated measures ANOVA tested the effects of the previously presented trial state on the percentages of states signalled on all omission trials and reported a significant effect of previous trial state on percentage of screen positions signalled on omission trials F (1, 17) = 0.30 p=0.034 (see table 8 for results summary). Interestingly though, while we recorded a significant effect of the previous trial on participants responses on omission trials indicating which taken in conjunction with reaction time data indicates that patterns were detected in sequences of stimuli they did not appear to use the transition probabilities in the way that predictive coding would predict. Instead of signalling the most likely state on each omission trial (optimal use), what participants actually seem to do could be explained by some combination of probability matching or pressing to signal the same screen position on each trial which could mean they might not have had an explicit prediction of the stimulus on omission trials. If we examine table 5, we can see evidence that supports this interpretation.

Table 6. Mean percentages of states signalled on omission trials relative to the previous state and 95% confidence intervals (lower and upper bound)-forwards and backwards transition matrices

State prior to omission trial (states D&H)	% Pressed 'A/E'	% Pressed 'B/F'	% Pressed 'C/G'	CI 'A/E'	CI 'B/F'	CI 'C/G'
A	22	48*	26	9, 34	38, 59	17, 35
B	19	35	41*	9, 29	21, 47	24, 58
C	33*	37	26	18, 46	27, 46	15, 36
E	27	29	32*	11, 42	11, 48	20, 47
F	30*	40	28	27, 49	19,61	24, 54
G	37	39*	21	17, 45	11, 41	10, 32

*Denotes predictable transition pairing

Table 5 above shows the percentage of presses for each screen position/state on omission trials. While we do see some evidence that participants are picking up the transition sequence probabilities they do not signal the most likely all the time. What appears to be occurring is that they either match the overall sequence of probabilities to some extent or some other perhaps some other strategy. If we look at A-omission trials the probabilities of state A staying in A was 11%, switching to B was 66% and to C also 11%, on B-omission trials the probabilities of staying in B are 11%, switching to A, 11% and to C 66% while in C-omission trials the probabilities of staying in C are 11%, switching to A, 66% and staying into C 11%. The same percentages exists in reverse matrix trials with the star indicating the transitions which have a 66% chance of occurring. Given these overall probabilities we can see that it appears that participants appear to be matching probabilities but it must be said not entirely. In forwards matrix omission trials participants generally press to signal the most likely state with presses distributed around the overall transition probabilities. For example in A-omission trials they press to signal the predictable B transition 48% of the time but also press to signal the other two possible states 22%(A) and 26% (B) of the time. This might be considered a closer result to probability matching than the optimal prescribed in predictive coding but still it is not entirely a match. Similar patterns exist in B-omission trials and C-omission trials with a noisier picture in backwards omission trails. In order to look at the where any significant differences lay in percentages pressed relative to previous trials in trial pairings we performed pairwise comparisons (Bonferroni corrected) on all transition pairings

Post hoc tests Forwards matrix transitions.

A-D (omission trials)

Analysis of percentages on A-omission trials reported significant differences between the predictable state B (M=48%, SD=15) and state A (M=22%, SD=18) with t (10) =-2.85, *p*=0.17. Significant differences were also recorded in percentages signalled between the predictable state B and state C (M=26, SD=13) with t (10) =3.26, p=0.08. No significant differences were recorded between states C and A with t (10) =4.90, p=6.34. This pattern of differences between states is not consistent with predictive coding in which percentages would have been nearly at 100% for the predictable pairing. Here we perhaps observe a somewhat noisy account of probability matching in which participants match the overall probabilities of transitions.

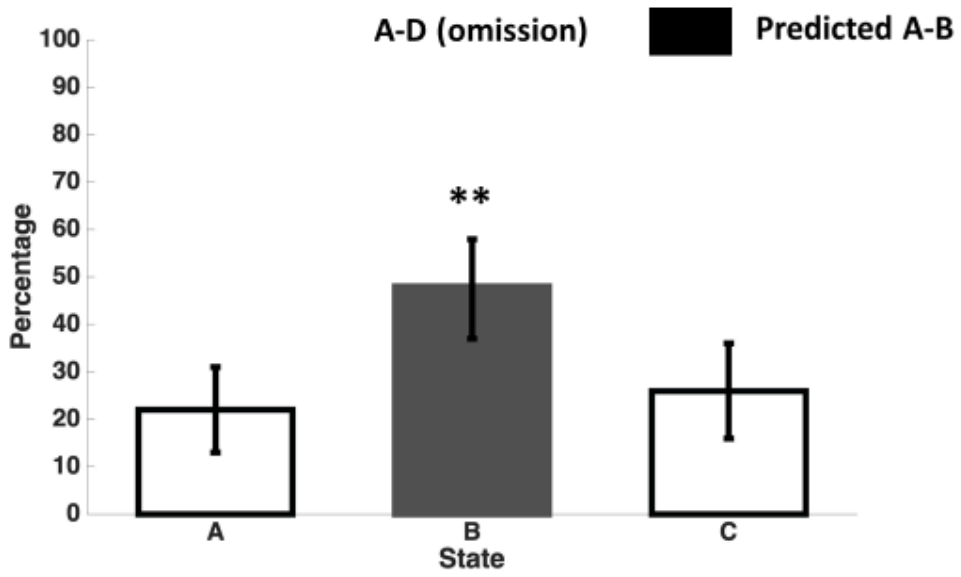


Figure 13. Percentage of states signalled during A-D (omission trials). Here participants did signal a significantly higher number of presses for the predictable state B transition in comparison to states A and C.

B-D (omission trials)

Paired samples t tests of percentages of states signalled on B-omission trials between state transitions reported no differences between the predictable state C (M=41%, SD=25) and state A (M=19%, SD=14) with t (10) =-1.97, *p*=0.76. No differences were recorded in percentages signalled between the predictable state C and state B (M=34, SD=19) with t (10) =0.51, *p*=0.621. No differences were recorded between states B and A with t (10) =2.16, *p*=0.55. Interestingly while post hoc tests did not detect significant differences between the percentages for the predicted state C and states A and B, state C had a higher mean percentage of presses than the other two states. Again, perhaps the best explanation is perhaps observe a somewhat noisy account of probability matching in which the most probable transition is selected more often but also that less predictable transitions also selected in a way quite close to the overall transition probabilities.

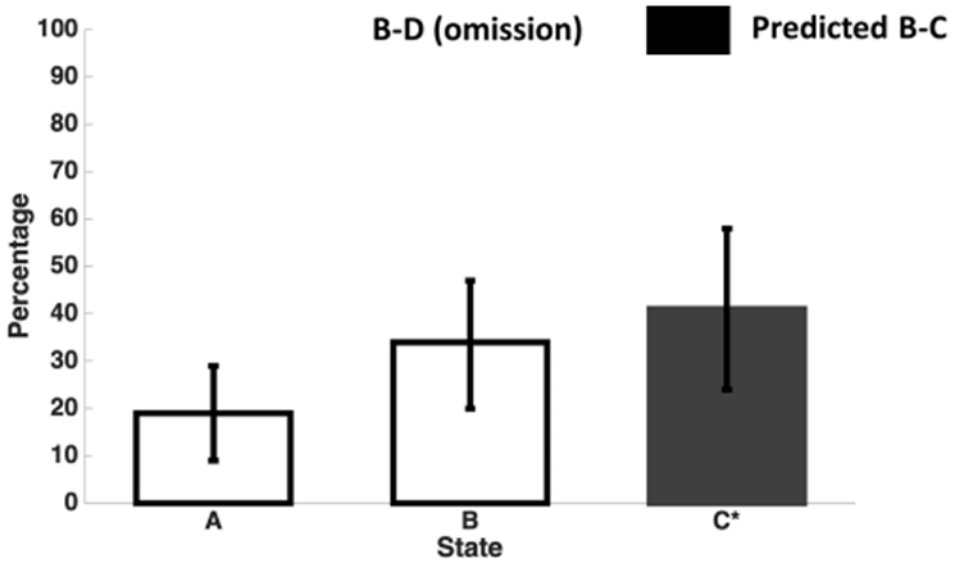


Figure 14. Percentage of states signalled during B-D (omission trials). Here participants again signalled a higher number of presses for the predictable state B transition although the difference was not significant.

C-D (omission trials).

Paired samples t tests of percentages of states signalled on C-omission trials between state transitions reported non significant differences between the predictable state A (M=32%, SD=21) and state B (M=0.36, SD=0.14) t(10)-0.43, *p*=0.67, state A and state C (M=26, SD .15), t(10)=0.61, *p*=0.552 and states B and C, t(10)=-1.54, *p*=0.154. This result is at odds with A and B to omission trials and there does not appear to be any explicit prediction on omission trials on this occasion.

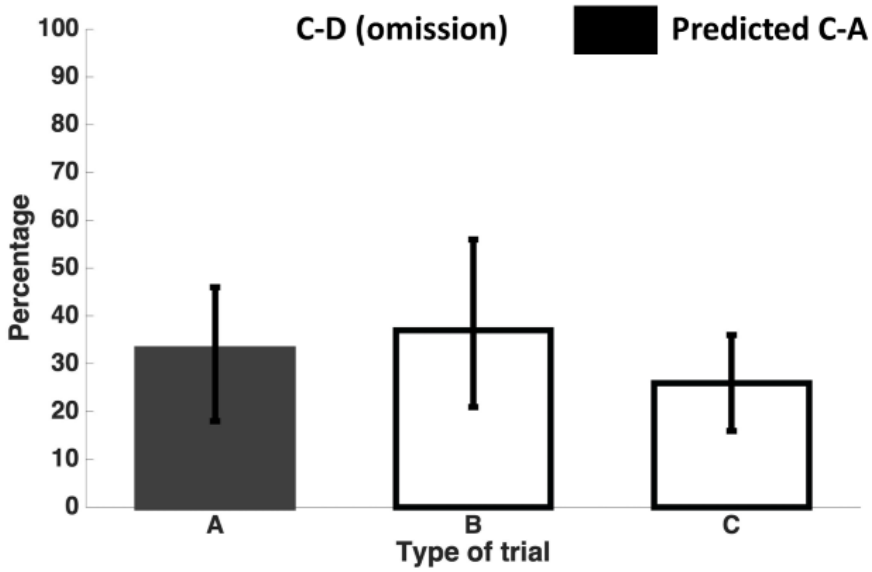


Figure 15 Percentage of states signalled during C-D (omission trials). On C-omission trials the overall picture was unclear and data noisy. Participants did not appear influenced by transition probabilities or the previous state.

Post hoc tests-Backwards matrix.

E-H (omission trials)

Paired samples t tests of percentages of states signalled on E-omission trials between state transitions reported non significant differences between the predictable state G (M=33, SD=20) and state E (M=27%, SD=23) with t(10)=-0.70, *p*=0.499, states G and F (M=29%, SD=27%), t(10)=0.68, *p*=0.509 and states E and F with t(10)=-0.24, *p*=0.814. One again, the data appears noisy. Participants did press to signal the predicted state more often but not by enough to draw any real conclusions and the result is not consistent with the optimal use of transition probabilities.

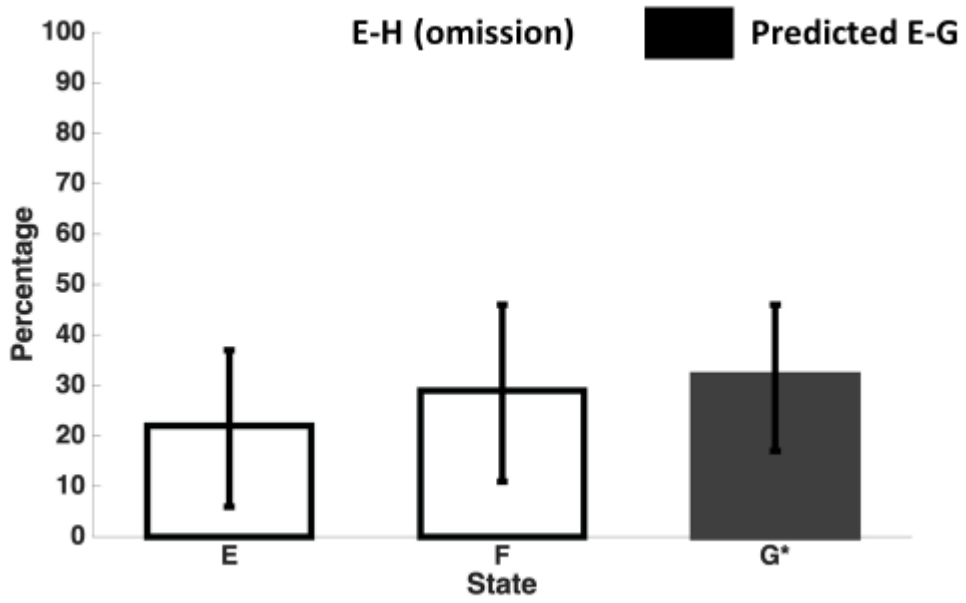


Figure 16 Percentage of states signalled during E-H (omission trials).

On E-omission trials the overall picture was again unclear and data noisy. Participants did press to signal the predicted G state more often but not by a significant amount.

F-H omissions

Paired samples t tests of percentages of states signalled on F-omission trials between state transitions reported non significant differences between the predictable state E (M=38%, SD=16) and state F (M=30%, SD=27) with t(10)=-1.35, *p*=0.207, states E and G (M=28%, SD=27%), t(10)=0.68, *p*=0.509 and states E and F (M=40%, SD-0.30) with t(10)=-0.38, *p*=0.707. The pattern of result here is again noisy and participants do not appear to be following the transition probabilities and respond approximately the same number of times to each screen position indicating no clear prediction,

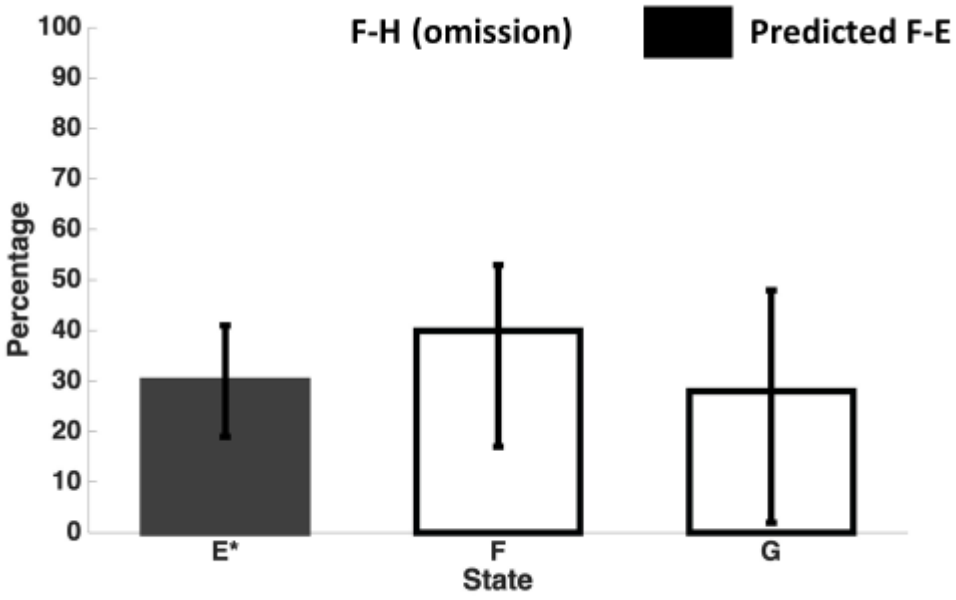


Figure 17. Percentage of states signalled during F-H (omission trials). On F-omission trials participants signalled roughly the same amount of times to each potential state.

G-H omissions

Paired samples t tests of percentages of states signalled on G-omission trials between state transitions reported non significant differences between the predictable state F (M=37%, SD=21) and state E (M=38%, SD=14) with t(10)=-0.13, *p*=0.897, states F and G (M=23%, SD=12%), t(10)=1.91, *p*=0.085 and interestedly we recorded significant differences between the non predictable states E and G with t(10)=-2.34, *p*=0.041. However, again here participants do not appear to be influenced by transition probabilities and it is unclear why they signal the unpredictable stimulus F state more often and may be somewhat of a random effect.

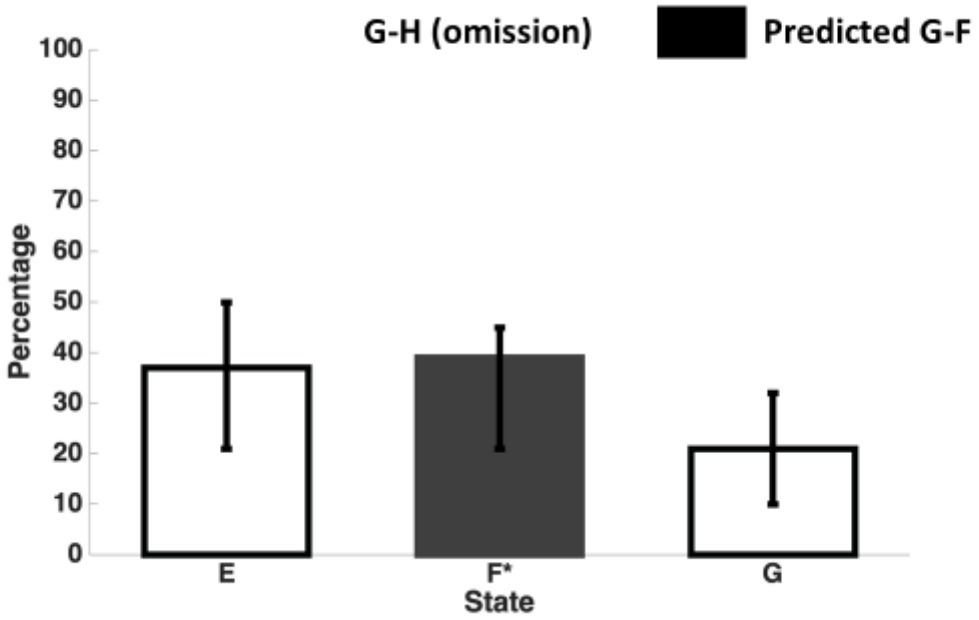


Figure 18. Percentage of states signalled during G-H (omission trials). On F-omission trials results indicated a significant differences between a non predictable

Reaction times-omission trials versus visible trials.

Reaction times were recorded in omission trials mainly to make a comparison with non omission, visible trials. Results reported a mean reaction time of 0.8651 (SD=0.19, CI [0.7851, 0.9242] contrast this with results from visible trials which over all reported mean reaction times of 0.4867 (SD=0.03, CI [0.4612, 0.5012]. This means that reaction times are approximately 50% faster in visible trials while this might have been expected it may provide some insight into the different strategies applied in omission trials as the longer reaction times may hint at the involvement of higher level decision making processes than used in visible transitions.

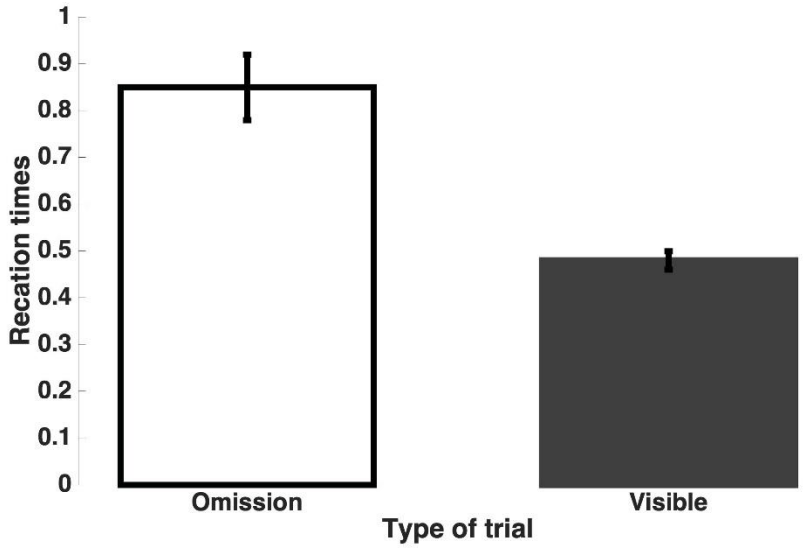


Figure 19 Reaction times-omission trials versus visible trials-Here as expected we observe faster reaction times in visible trials to our omission trials

5.12 Discussion and conclusion.

Visual information is variable and uncertain presenting a complex problem for perceptual systems. In order to deal with uncertainty it is proposed the brain attempts to model and predict its incoming inputs (Garrido, Kilner, Stephan, & Friston, 2009; Kirchhoff, Parr, Palacios, Friston, & Kiverstein, 2018; Summerfield & de Lange, 2014). In predictive coding, this ability rests on the existence of a series of hierarchical models containing representations of the statistical regularities of the external world that the brain constantly refines and updates (Friston, 2010; Parr et al., 2018; Rao & Ballard, 1999). Perhaps the richest source of temporal regularity available to the brain are the conditionally probabilistic relationships that exist in sequences of information created by the interactions and behaviours of stimuli in the world. By extracting and applying this sequential information, it can in theory provide a rich source of predictive information that can be used to improve the effectiveness of behaviours. However, while sequential regularities are replete with predictive information, on occasion they are quite subtle in comparison to other types of temporal regularities and it is unclear what type of temporal regularities and relationships the brain actually uses to make predictions.

It is possible that the brain uses more simple and explicit types of temporal regularities to make predictions. These type of regularities include cues or repetitions of events offering a simpler if less rich source of predictive information. Furthermore, it is also unclear as to how any potential top down input predictive information based on sequences effects how people actually respond or make decisions in a way that makes behaviour more effective. Previous predictive coding literature has seldom insert complex sequences into their stimulus sequences and are not able to test such ideas (Kok, Jehee, & de Lange, 2012b; Kok & Turk-Browne, 2018; Summerfield et al., 2008) and a number of studies that have used conditional sequences of stimuli have shown that they do not manifest in the decision making process in a way that optimises behaviour (Jones & Pashler, 2007; Maljkovic & Nakayama, 1994) counter to predictive coding and that repetition or other sources of temporal information may be applied.

In the current chapter, we set out with a clear theoretical motivation that aimed to compare two competing accounts of how the brain uses temporal regularities. One was the predictive coding account (Friston, 2010) and one based on findings from Maljkovic & Nakayama (1994) was the short term memory or priming account. Predictive coding states that the brain is constantly making and changing its predictions as event and sequences unfold in the world, On the other hand, in priming, the only thing that matters is that events occur more often (Goolsby & Suzuki, 2001; Henson, 2003; Kristjánsson, 2006; Maljkovic & Nakayama, 1994; Olivers & Meeter, 2008; Wiggs & Martin, 1998). The key manipulation in the current chapter was inserting a sequence of stimuli in our experiment that manipulated conditional predictability but key to the priming account held constant the total number of screen positions. This can be thought of as a similar type of experiment to Maljkovic & Nakayama

(1994) in which they also manipulated the probability or expectancy of stimulus change but held the actual number of switches between colours constant but here instead of just manipulating change or switch we inserted a sequence that made specific stimuli positions more or less predictable. This was achieved by using two Markov chains into our stimulus blocks we called forwards and backwards matrix. Here, we found a different outcome to Maljkovic & Nakayama (1994) who had reported slower reaction times to more predictable stimuli. Reaction times strongly supported the idea that sequences of information can be used by the brain to improve behaviour with a significant effect of predictability on reaction times.

Reaction times support predictive coding

Faster reaction time for predictable screen positions can be interpreted as the brain using the most likely stimulus transitions to reduce overall prediction error thus leading to faster response times as in predictive coding (Summerfield & Koechlin, 2008). Further validity to the predictive coding account of the brains ability to apply complex sequential information when making predictions comes from the significant interaction effect we found between transitions matrix direction and reaction times. This result can also be interpreted within the predictive coding account of integration as that the brain had learned the transition probabilities during the forwards matrix trials and had begun to relearn them during the backwards matrix trials. However, while the results from reaction time data are consistent with the predictive coding account integration and show that conditional relationships can improve behaviour one must question why our results were so different from Maljkovic & Nakayama (1994) which it must be said was a very well controlled and rigorous study .

One explanation might be the idea that when there is a specific predictable outcome people do use sequential information to predict and that in such occasions the frequency of transitions is less informative. Another potential explanation is that Maljkovic & Nakayama (1994) task could be thought of as more of a slightly higher level explicit decision making task as opposed to an implicit faster visual task judgment. If we look at stimulus timings our stimulus was presented for 200 ms while Maljkovic & Nakayama (1994) stimulus remained on the screen for 2.5 seconds. It may have been possible that under situations that require an explicit decision the brain cannot access the transition probabilities so readily as a faster view-respond experiment which may be learned on a more implicit basis. This idea would be worthy of future experimentation that presents our stimulus at different time lengths to see if stimulus exposure time is a factor. Another, possibility is that Maljkovic & Nakayama (1994) stimulus expectancy versus priming experiment part of the paper only tested three participants (two of those were Maljkovic & Nakayama). As we have found previously how people use past information exhibits high inter-individual variability potentially subject to individual task goals (Kirchhoff et al., 2018; Parr et al., 2018) and variances in learning the behaviour

of the distal stimulus that makes them non optimal in predictive coding terms. It would therefore, be interesting to replicate Maljkovic & Nakayama (1994) with a larger and more naïve participant set and change stimulus presentation timings. However, it is possible that participants displayed a non optimal decision making strategy linked to exploring changes in stimulus sequences that may have been more evident in our omission trial data.

Potential non predictive coding decision strategies employed on omission trials.

While reaction time data supported the predictive coding account our omission trial data somewhat muddies the waters in terms of how the use of conditional relationships and sequences of stimuli translate into making a decision. Furthermore, omission trial data also asks questions about the optimal use of information in sequences of stimuli or more specifically what is optimal. Here, there is little doubt that participants were able to learn the transition probabilities but as in Jones & Pashler (2007) it asks question as to how people use them in terms of decisions. As we have seen reaction time data shows that participants had learnt the transition probabilities and had even been able to re learn the transitions when we changed the direction of the transition matrices. Furthermore, not only were participants able to learn the transitions they could actively use them to increase the effectiveness of behaviours in faster reaction times. This made our result on omission trials all the more intriguing. Here, although it looks as though there is a response pattern supporting that participants were using the transition probabilities to some extent on omission trials but not in a way consistent with reaction time data or predictive coding theory. What actually seems to have happened is that we observed some evidence for probability matching and other signs suggesting that people did not apply them as an explicit prediction.

Looking at percentages signalled on omission trial relative to previous trials it can be considered that on some occasions participants exhibited probability matching in their responses. Probability matching is when instead of choosing the optimal response in a stochastic context in terms of error/reward on the current response which should always be the most probable choice, instead match the overall probabilities of the observed sequence (Newell et al., 2013). Interestingly, this strategy is also found in non-humans species (Li & Dudman, 2013; Niv, Joel, Meilijson, & Ruppin, 2002; Robbins & Warner, 1973) and is actually found in nature more commonly than what might be considered optimal strategies (Koehler & James, 2010). We found that, to some degree, participants when faced with stimulus transitions that had percentages of switching to another state of 66%, 11% and 11% (with omission trials making up the other 11%) responses did match these probabilities in their decisions albeit less than perfectly. Such a phenomena is well recorded in decision making literature (Koehler & James, 2010) and raises questions about the overall goals of the use of

sequential information and in fact what is optimal in terms of reducing error which may still tie in with predictive coding but in a more complex way than simply optimizing error on current trials.

If we think about the probabilities of how events occur in the world it is true to say that while some events are more probable given a previous event but also that these probabilities can change. Now, if we think how we might learn about changes in probabilistic relationships a possible reason for probability matching becomes apparent. If we always make the same most probable decision we would not explore how events in world might be changing and potentially miss some important new predictive relationship. It may be that in order to maximise rewards in the long term people apply some focus to learning about how events might be changing to maximize reward in the longer term by experimenting with different choices to see if things are changing. This idea might be given credence given that in our experimental blocks probabilities were changing during blocks as the matrices reversed meaning that the brain might have detected an unstable series of sequence probabilities and investigated this change further by matching probabilities. Considering that predictive coding states that the brain is actively seeking out information from the world this may in fact be an extension of that process and lead to reduced prediction error in decision making over longer time scales. However, it must also be said that this does not appear to be the only influencing factor and it may be that internalising sequential information does not favour forward relationships as seen in Jones & Pashler (2007)

In Jones & Pashler (2007), they found that even though their stimulus sequence technically favoured relationships this did not manifest in participants explicit decisions. This can in some ways be considered similar to our results. We also found that participants could learn the transition probabilities and could apply them in a reaction time task but this did not seem to carry over how when participants were faced with a situation in which they had to make a more explicit prediction about the next trial. This may be explained by the idea that temporal regularities simply form associations like in cue or priming experiments that need not necessarily be forwards in direction and that prediction, is not as Jones & Pashler (2007) state, the overriding goal of the brain and it simply associates temporally adjacent events.

Study limitations.

There are also a number of explanations for our somewhat inconsistent results that may be explained by the limitations in our study. One is that task had no real value to correctly predicting the stimulus on omission trials. It has been previously suggested that dopamine and reward pathways may modulate how we form judgments and make decisions (Montague, Hyman, & Cohen, 2004). It may have been possible that by providing some monetary or other reward for ‘correct choices’ we may have encouraged participants to apply the transition probabilities on omission trials in a more

traditionally optimal way. Another potential factor that may have been caused by limitations in our analysis techniques is that we did not analyse how responses altered over the course of trials. This critique while still applying to reaction time data is more strongly directed to omission trial analysis. It is possible that decision strategies changed over the course of trials. By only analysing the overall strategy we may have missed switches in how the brain applied the transition probabilities by only looking at overall responses. It may have been that participants had no explicit prediction at the start of blocks but did form predictions over the course of trials or matched probabilities and then changed strategy. Predictive coding states that the brain constantly updates its internal models as new information becomes available but we did not assess this aspect of the theory aside from assessing that behaviour had changed during backwards matrix trials as opposed to the time course of responses. In future experiments this analysis might provide further explanations of our data.

Conclusion.

There are a number of key aspects presents in results from chapter 5 that provide insight into the predictive strategies used by the brain. One is that the brain can learn and apply conditional relationships present in sequences of stimuli behaviourally and is not restricted to learning cues or frequencies when making predictions. We found that reaction times for predictable stimuli were faster than those for non predictable stimuli. This is consistent with the ideas of predictive coding, in which favouring predictable stimuli decreases prediction error making behavioural responses more effective and inconsistent with ideas of priming or memory effects. However, our omission trial analysis raised questions about the simple interpretation of reaction times from predictive coding experiments. In this analysis we found that the relationship between improved reaction times and more explicit decisions was not straightforward. According to predictive coding, sequence based predictions arising from top down inputs should always advantage the predictable stimulus in an attempt to reduce prediction error. We did not find this outcome. Instead we found a mixed picture that reflected a combination of non optimal decision making strategies in which participants appeared to either match the overall transition probabilities or did not have an explicit prediction. It may be possible that the matching probabilities is still consistent with the overall aim of predictive coding but aims to optimise prediction error on a longer term by exploring how new decisions match with incoming information that could have changed. However, it is also possible that sequential relationships do not form explicitly forward temporal predictions and may simply link together temporally adjacent events. In summary we have support for predictive coding in terms of reaction times but like other studies we cannot draw a firm conclusion of how decreased reaction times based on learnt sequential information translate into the decision making process.

Chapter 6. General discussion, contribution and concluding remarks

The aim of this PhD was to assess the validity of the application of predictive coding in time and provide new theoretical and experimental ideas to guide research in this area. To deal with variable inputs in isolation from the world a number of current predictive coding models assume that the brain contains models of the environment and adapts the representation of the world within its models based on precision weighted prediction errors (Friston, 2010; Spratling, 2015). These ideas include, the learning of conditional probabilities present in the relationships in events and the behaviour of stimuli in the environment and weighting prediction errors based on the reliability of sensory information and previous predictions (Friston, 2017; Thornton, 2014). In particular, we aimed to examine the factors which examine the precision weighting of prediction errors and the ability of the visual system to extract and the ability of the brain to use complex conditional relationships present in sequences of stimuli. Using a number of computational and mathematical frameworks to provide clear and testable hypotheses, we conducted a number of experiments detailed in chapters 3, 4 & 5 and obtained the following key experimental results, methodological and theoretical advances.

6.1 Key experimental findings.

Chapter three

- To minimize the effects of uncertainty, the visual system integrates past and current sensory information adaptively. Integration is tuned to the variability of sensory measurements (proximal variance) and the behaviour of the stimuli of interest (distal variance). When proximal variance was higher relative to distal variance participants perceptual estimates were closer to previous stimulus values and when distal variance was higher relative to proximal variance estimates were closer to current stimulus values.
- The related integrative visual phenomena termed serial dependence is adaptive. In contrast to previous characterizations of the phenomena (Jennifer E Corbett et al., 2011; Kiyonaga et al., 2017; Liberman et al., 2014, 2016), it is not pervasive functioning at a fixed level and instead adapts to level of variability in the behaviour of stimuli (distal variance) and the reliability of the observation of the stimulus in question (proximal variance) more consistent with ideas from Bayesian visou-motor literature (Denève, Duhamel, & Pouget, 2007; Knill & Pouget, 2007; Wolpert, Ghahramani, & Jordan, 1995; D. M. Wolpert & Flanagan, 2001)
- Temporal integration strongly favours the current stimulus input over previous information. Despite designing a stimulus that was extremely hard to see it was very difficult to observe the effects of past stimulus history in participants estimates until we greatly reduced distal variance.

Chapter 4.

- Correction of perceptual error is adaptively modulated by the variability of sensory measurements (proximal variance) and the behaviour of the stimuli of interest (distal variance). When proximal variance was higher relative to distal variance, participants estimates corrected faster when stimulus values changed and when distal variance was higher relative to proximal variance estimates corrected slower to changes in stimulus values. Again, a result consistent with from Bayesian visou-motor literature (Burge et al., 2008, 2010)
- The way participants integrate information over time exhibits high individual variability. Some of this variability can be explained by the different levels of proximal variability in participant's observations but not all.
- Again, we found that temporal integration strongly favours the current stimulus input over previous information. Although, we observed significant differences in Kalman gain which acted as our proxy measure for error correction, Kalman gains did not drop below the 0.7 level in any condition indicating a much higher effect of the current stimulus value on estimates than past stimulus values.

Chapter 5.

- The visual system can extract and use conditional probabilities in the way stimuli sequentially change over time to improve behaviour consistent with current predictive coding theory (Friston, 2010; Spratling, 2015). We found that most probable or predictable transitions, elicited faster reaction times than less probable transitions.
- When the sequence of stimuli is broken by omitted stimulus values, participants appear not to use the probabilities in the optimal way prescribed by predictive coding.
- Our experiment showed that participant's responses on current trials were the result of a combination of a non optimal decision making strategy, termed probability matching, or not applying a clear prediction to the omitted trial. This finding raises questions about the translation of predictive coding to decision making.

6.2 Important methodological advancements.

Chapter 3.

- Our proximal variance calibration experiment generally provides a reliable means to quantify visual variability. While data from this experiment may include variability from other neural systems, overall this method provided an accurate means to assess individual levels of visual variability and figures to provide a group level. Given that sensory uncertainty is a prominent area of vision science (Aitchison & Lengyel, 2017; Albright, 2012; Clark, 2013; Friston, 2010; Kok, De Lange, Kok, & De Lange, 2014; Kwisthout & Van Rooij, 2017) this technique provides a simple but effective way to ascertain the levels of uncertainty in visual perception for computational purposes.
- The Kalman filter is an excellent theoretical framework for guiding future temporal predictive coding experimental designs. Its concepts of proximal and distal variance, touch on something often missed in predictive coding studies and are simple to manipulate with existing experimental designs.
- Kalman gain is an excellent predictor of the relative influences of past and current information of current estimates. In both experiments in chapter 3, Kalman gain predicted the level of serial dependency we would observe almost perfectly.

Chapter 4

- Here, we draw attention to the replication of results from the proximal variance calibration experiment performed in chapter 3. Results in chapter 4 showed that the method is robust and reliable in a totally new participant set.
- The step response function provides an effective means to assess how perceptual error is reduced over time. Most experimental designs that examine temporal integration change stimulus values at random over trials. Changing values on every trial makes it hard to assess how error reduces over time to zero. As this is a key aspect of predictive coding (i.e. the actual aim of predictive coding) this step based design, which steps then stabilizes, is ideal for analysing this aspect of predictive coding.
- Kalman gain is a reliable and easy to conceptualize proxy measure of perceptual error correction. Kalman filter theory states that higher Kalman gains lead to faster correction of estimation and lower Kalman gains lead to slower correction. We found this model to be very accurate and as Kalman gain is represented in one single number is straightforward to understand

Chapter 5

- Markov chains are an effective means to introduce predictability into sequences of stimuli that change over time in predictive coding experiments. Previous studies in the field have often not contained any real level of predictability but this is something that the insertion of Markov transition probabilities into trial sequences can rectify.

6.4 General comment on the use of ideas from control theory

- We also mention the general success in incorporating ideas from control theory and signal processing into predictive coding studies. These areas are increasingly turned to by Neuroscientists (Burge et al., 2010; Cicchini, Anobile, & Burr,2008; Faisal, Selen, & Wolpert, 2008; Todorov, 2004; Wolpert, Ghahramani, & Jordan, 1995) as they provide huge expertise in many problems which are in essence the same as in vision science and motor control. This is especially true if one wants to understand issues that have been studied for many years in signal processing such as error weighting and correction as shown in the current thesis.

6.5 Theoretical implications.

At the heart of predictive coding lies the notion of a predictive system of models in probabilistic form that updates predictions based on precision weighed prediction errors (Clark, 2013;Friston, 2010). While there have long existed models which explain how this system might actually work in other areas of visual processing (Huang & Rao, 2008; Rao & Ballard, 1999), previously no model dedicated to predictive coding in time existed. However, based on our findings the Kalman filter is a at some levels, a suitable candidate with support for the models account of the weighting precision weighting of prediction errors and indirectly the internal modelling of conditional probabilities albeit with a number of limitations.

6.6 Precision weighting of prediction errors.

Our results show strong behavioural correlates of the Kalman filter account of precision weighting consistent with an existing body of existing Bayesian visou-motor work (Burge, Ernst, & Banks, 2008; Burge et al., 2010; Knill, 2007; Kwon, Tadin, & Knill, 2015; Todorov, 2004) and predictive coding (Friston, 2010). In our view, this is an important step on the path to establishing the ability of the visual system to interpret information taking into account the estimated reliability of sensory information and uniting findings from distinct if related research areas. Importantly, while we could not study the neural signatures of prediction error we were limited to examining behavioural signatures. However, all of our results were consistent with the predictive coding account of precision weighting. Our results support the idea that less weight is attached to less reliable information and more weight to more reliable information. In computational terms this weighting strategy represents a

method for the brain to optimally balance past and current information to account for variability in its inputs (Denève et al., 2007b). Our experiments in establishing support for this balance shows the need for more experimental designs that test the way the reliability of sensory information affects the way errors are weighted. This would be especially interesting in terms of Neuroimaging studies that have studied prediction errors which generally present clearly visible stimuli with the aim of eliciting the strongest possible imaging signal to observe any changes to such signals. Basing experiments of the principles of the Kalman filter and manipulating proximal and distal variance and observing neural activity could be a productive means of advancement for such studies.

6.7 The extraction and use of sequential information in making predictions.

Results from chapter 5 provide a mixed picture about the use of conditional probabilities present in the sequences of stimuli in making predictions. On one hand we show that such information can be integrated to improve behaviours in terms of reaction times but exactly the relationship between predictive coding and decision making is unclear it appears that in our experiment participants did not use the transition probabilities in decision in the way predictive coding would predict and warrants further investigation. In terms of methods though our study was very successful.

The use of Markov chains is also a key aspect of our study that we take from other areas of the literature and apply to predictive coding. Markov chains and related concepts that we have previously not discussed such as Markov blankets are used in a number of computational predictive models (see Kirchhoff, Parr, Palacios, Friston, & Kiverstein, 2018 for an introduction). However, such models are extremely complex and difficult to test in simple terms. Here, we show that Markov chains can be a productive model for providing behavioural support for such ideas. The most obvious area of support was by explicitly inserting Markov transition probabilities in chapter 5 and the second was more indirect which provided more support the idea that the Kalman filter is a useful predictive coding model in the temporal domain.

In chapter 3's experiment two we inserted a one dimensional random walk into the stimulus sequence with the primary aim of reducing distal variance. A one dimensional Gaussian random walk will inevitably reduce variability in comparison to a random stimuli but importantly they are considered a special category of a Markov processes which the Kalman filter requires the underlying stimulus behaviour in question to follow. Because in this experiment participants only integrated past information under conditions of uncertainty when we inserted the Gaussian random walk this result also provided additional indirect support for both the idea that the brain relies on conditional probabilities in conditions of uncertainty and the use of the Kalman filter as a predictive coding model in the temporal domain.

6.8 Limitations of the thesis.

Overall the theory and experimental designs contained with the current thesis were successful but of course this does not mean that they are not without their limitations. One potential limitation relates to exactly how much of our measured quantities, be they proximal variance or serial dependence, came from the activity of the visual system alone. For example, in our proximal variance calibration the task involved judging the change in two stimulus orientations separated in time. While the delay in presentation was short this still involved some working memory component. Like any neural function working memory will collect variability from factors such as neural noise meaning that our proximal variance data may be inflated.

The same critique can also be applied to our serial dependence data and the way we have interpreted our findings. Serial dependence has in the main been considered to be a visual phenomenon (Fischer & Whitney, 2014; Kiyonaga et al., 2017; Liberman et al., 2014). However, recently this view has been challenged. Some very recent studies have stated that at the early stages of visual processing serial dependence is not present and in fact that results from post perceptual memory processes This standpoint is based on findings that behavioural responses made immediately observing a stimulus in fact exhibit a repulsion effect as opposed to an attractive bias as in serial dependence (Bliss, Sun, & D' Esposito, 2017; Fritsche, Mostert, & de Lange, 2017). Only when the time delay between stimulus and response were lengthened was any serial dependence observed. The interpretation applied to these findings is that the first function of visual processing is to differentiate stimuli which would provide more accurate perceptions and only later is it beneficial to apply serial dependence because working memory is vulnerable to distraction and more noisy making smoothing information a good strategy (Fritsche, Mostert, & de Lange, 2017. It is fair to say that our stimuli and judgment responses were separate in time by 3-4 seconds meaning that we may also have recorded post perceptual serial dependence. Furthermore, we also observed a repulsion effect in our serial dependence experiment and an overshoot in our step design experiments in some instance that may be consistent with this idea. We surmised this was down to an incorrect model of the behaviour of stimuli but timing between stimulus and judgment may have played some role. Indeed further experiments involving a number of time spans between stimulus and judgment this would be an interesting follow up experiment given the debate between the two serial dependence camps.

A further limitation of the current thesis was highlighted by the individual differences in integration strategy we observed in our experiments. In some cases, participants appeared to nearly always integrate past information in current estimates while in other cases the opposite was found. One reason was the level of proximal variance in individual participants. We did find a moderate negative correlation between proximal variance and Kalman gain indicating that proximal variance could explain some individual variability. However, it appears that proximal variance could not account for

individual differences entirely. One possible explanatory factor is the task prior that participants applied. In some predictive coding model there exists within the internal models not only a probabilistic model of the behaviour of the external stimulus but also a model of task structure (K. Friston, 2010). In more simple terms what the brain considers the goal of the task to be. If individual participants had different models of task structure then this could explain some of the individual variability we observed. For example, for some the task structure might have been to reduced noise by temporally smoothing over values. These participants would therefore be more serially dependent and slower to correct to change. Alternatively, some participant's task structure might have been to differentiate between stimulus values. If this was the case, then this might explain the repulsion effects we observed. Importantly, this highlights a large flaw in the Kalman filter. The Kalman filter is a model from signal processing that frankly does not care about task demands. All it does is optimize mean squared error under certain assumptions. In humans, the aims could be much different in certain people and the Kalman filter is not capable of testing this idea. Future experiments could further investigate the role of task demands by manipulating the instructions to participants and attempt to model outcomes under models such as free energy (Friston, 2002) which does include task priors in its calculations. However, by doing you would remove the one of the main attractions of using the Kalman filter as a temporal predictive coding model-its simplicity and ease of testing.

Our last limitation, although relating to our study directly also provides a thought provoking limitation of predictive coding in general. Predictive coding, in whatever its guise places a very strong emphasis on perception as a prediction based on previous experience. If one was to read the literature, you could be forgiven for thinking that what we observe at any moment is nearly entirely the 'prediction' generated by the brains internal models. However, generally we found this was not strictly the case. In our experiments that modelled responses with the Kalman filter, we found the dominant factor to be the current stimulus. In reality, we found that it was very hard to induce the visual system to integrate past information. Only in very uncertain circumstances was the influence of past information observable. Of course, you could interpret this as the success and speed of the predictive coding system in which internal models are nearly almost entirely accurate to the world at all times. Alternatively, you could interpret this as the visual perception being a more bottom up process only requiring predictive information in times of great uncertainty. Furthermore, in experiment 5, we found that the participants did not apply the use of probabilistic information in the optimal, predictive coding way and instead applied other strategies. Both sets of findings ask questions about the way the brain uses past information and the way we make predictions.

6.9 Concluding remarks.

To conclude, this thesis aimed at establishing the validity of the predictive coding account of temporal integration. To do this, we used a variety of behavioural measures including serial dependence and analysing participants step responses and reaction times in conjunction with the computational and mathematical frameworks provided the Kalman filter and Markov chains. Our findings relating to serial dependence and the way participants corrected error, both largely support the predictive coding account of temporal integration. Also, the Kalman filter framework provided a useful model of predictive coding in time and provided solid explanations of predictive coding's ideas about the precision weighting of prediction errors. In terms of testing how people use probabilistic sequential information, our use of Markov chains proved suitable for testing how people integrate sequential information to make predictions over time. Overall, we support the predictive coding account but raise questions about the application of predictive information in such an all explaining manner as in some accounts of predictive coding and also question how predictive coding translates into decision making processes.

7. References

Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/J.CONB.2017.08.010

Alais, D., Leung, J., & Van Der Burg, E. (2017). Linear summation of repulsive and attractive serial dependencies: orientation and motion dependencies sum in motion perception. *J. Neurosci*, *10*, 4601–4615. https://doi.org/10.1523/JNEUROSCI.4601-15.2017

Albrecht, A. R., & Scholl, B. J. (2010). The Nature of Perceptual Averaging: Automaticity, Selectivity, and Simultaneity. *Journal of Vision*, *10*(7), 1252–1252. https://doi.org/10.1167/10.7.1252

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810–815. https://doi.org/10.3758/s13414-012-0293-0

Albright, T. D. (2012). On the Perception of Probable Things: Neural Substrates of Associative Memory, Imagery, and Perception. *Neuron*, *74*(2), 227–245. https://doi.org/10.1016/j.neuron.2012.04.001

Anderson, N. H. (1967). *Application of a weighted average model to a psychophysical averaging task 1*. *Psychon. Sci* (Vol. 8). Retrieved from https://link.springer.com/content/pdf/10.3758%2FBF03331634.pdf

Arathorn, D. W., Stevenson, S. B., Yang, Q., Tiruveedhula, P., & Roorda, A. (2013). How the unstable eye sees a stable and moving world. *Journal of Vision*, *13*(10), 22–22. https://doi.org/10.1167/13.10.22

Ariely, D. (2001). *SEEING SETS: Representation by Statistical Properties*. *PSYCHOLOGICAL SCIENCE Research Article* (Vol. 12). Retrieved from http://journals.sagepub.com/doi/pdf/10.1111/1467-9280.00327

Auksztulewicz, R., & Friston, K. (2016a). Repetition suppression and its contextual determinants in predictive coding. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(5), 358–366. https://doi.org/10.1038/nrn1888

Bachmann, T. (1997). Visibility of Brief Images: The Dual-Process Approach. *CONSCIOUSNESS AND COGNITION*, *6*, 491–518. Retrieved from https://pdfs.semanticscholar.org/d5f1/9331b8ff25a0658d767ad25cb38eed803b5a.pdf

Baddeley, R. J., Ingram, H. A., & Miall, R. C. (2003). System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(7), 3066–3075. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12684493

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*(04). https://doi.org/10.1017/S0140525X01000024

Barnes, G. R., & Asselman, P. T. (1991). The mechanism of prediction in human smooth pursuit eye movements. *The Journal of Physiology*, *439*, 439. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1180117/

Bauer, B. (2017). Perceptual averaging of line length: Effects of concurrent digit memory load. *Attention, Perception, & Psychophysics*, *79*(8), 2510–2522. https://doi.org/10.3758/s13414-017-1388-4

Becker, S. I. (2008). Can Intertrial Effects of Features and Dimensions Be Explained by a Single Theory? https://doi.org/10.1037/a0011386

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23–23. https://doi.org/10.1167/9.5.23

Bell, I. R., Wyatt, J. K., Bootzin, R. R., & Schwartz, G. E. (1996). Slowed reaction time performance on a divided attention task in elderly with environmental chemical odor intolerance. *The International Journal of Neuroscience*, *84*(1–4), 127–134. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8707474

Bergen, R. S. van, & Jehee, J. F. M. (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *BioRxiv*, 671958. https://doi.org/10.1101/671958

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-4286-2

Berniker, M., & Kording, K. (2011). Bayesian approaches to sensory integration for motor control. *Ltd. WIREs Cogn Sci*, *2*, 419–428. https://doi.org/10.1002/wcs.125

Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5411–

5416. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9144251

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 14739. https://doi.org/10.1038/s41598-017-15199-7

Braun, D., Boman, D., & Hotson, J. (1996). Anticipatory smooth eye movements and predictive pursuit after unilateral lesions in human brain. *Experimental Brain Research*, *110*(1), 111–116. https://doi.org/10.1007/BF00241380

Breitmeyer, B. G., & Ganz, L. (1976). Implications of Sustained and Transient Channels for Theories of Visual Pattern Masking, Saccadic Suppression, and Information Processing, *83*(1). Retrieved from http://psycnet.apa.org/fulltext/1979-20169-001.pdf

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–1595. https://doi.org/10.3758/BF03212157

Breitmeyer, B. G., Rudd, M., & Dunn, K. (1981). Metacontrast investigations of sustained–transient channel inhibitory interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 770–779. https://doi.org/10.1037/0096-1523.7.4.770

Brockmole, J. R., Wang, F., & Irwin, D. E. (1985). Temporal Integration Between Visual Images and Visual Percepts. *Finke & Pinker*. https://doi.org/10.1037//0096-1523.28.2.315

Brostek, L., Eggert, T., & Glasauer, S. (2017). Gain Control in Predictive Smooth Pursuit Eye Movements: Evidence for an Acceleration-Based Predictive Mechanism. *ENeuro*, *4*(3). https://doi.org/10.1523/ENEURO.0343-16.2017

Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, *8*(4), 20.1-19. https://doi.org/10.1167/8.4.20

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(22), 7714–7721. https://doi.org/10.1523/JNEUROSCI.6427-09.2010

Burr, D. (2005). Vision: In the Blink of an Eye. *Current Biology*, *15*(14), R554–R556. https://doi.org/10.1016/J.CUB.2005.07.007

Burr, D., & Cicchini, G. M. (2014). Vision: Efficient Adaptive Coding. *Current Biology*, *24*(22), R1096–R1098. https://doi.org/10.1016/j.cub.2014.10.002

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551–566. https://doi.org/10.1113/jphysiol.1968.sp008574

Carl, J. R., & Gellman, R. S. (1987). Human smooth pursuit: stimulus-dependent responses. *Journal of Neurophysiology*, *57*(5), 1446–1463. https://doi.org/10.1152/jn.1987.57.5.1446

Casini, L., Mckay Illari, P., Russo, F., & Williamson, J. (2011). Models for Prediction, Explanation and Control: Recursive Bayesian Networks. *BIBLID*, *26*, 495–4548. Retrieved from https://blogs.kent.ac.uk/jonw/files/2015/03/rbns2011.pdf

Chen, S. Y. (2012). Kalman Filter for Robot Vision: A Survey. *IEEE Transactions on Industrial Electronics*, *59*(11), 4409–4420. https://doi.org/10.1109/TIE.2011.2162714

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception and Psychophysics*. https://doi.org/10.3758/BF03195009

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Perception, & Psychophysics*, *72*(1), 86–99. https://doi.org/10.3758/APP.72.1.86

Chopin, A., & Mamassian, P. (2012). Predictive Properties of Visual Adaptation. *Current Biology*, *22*(7), 622–626. https://doi.org/10.1016/J.CUB.2012.02.021

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M. (2003). Scene Perception and Memory. *Psychology of Learning and Motivation*, *42*, 79–108. https://doi.org/10.1016/S0079-7421(03)01003-X

Cicchini, G. M., Anobile, G., & Burr, D. C. (2011). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. https://doi.org/10.1073/pnas.1402785111

Cicchini, G. M., Anobile, G., Burr, D. C., Agrillo, C., Bisazza, A., Izard, V., … Tibber, M. S. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*, 12536. https://doi.org/10.1038/ncomms12536

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. https://doi.org/10.1167/17.14.6

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences (2013) 36, 181–253*, 181–253. https://doi.org/10.1017/S0140525X12000477

Cluff, T., Crevecoeur, F., & Scott, S. H. (2015). A perspective on multisensory integration and rapid

perturbation responses. *Vision Research*, *110*(Pt B), 215–222. https://doi.org/10.1016/j.visres.2014.06.011

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating Stable Representations: Serial Dependence in Vision. *PLoS ONE*, *6*(1). https://doi.org/10.1371/

Corbett, J. E., & Melcher, D. (2014a). Characterizing ensemble statistics: mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics*, *76*(3), 746–758. https://doi.org/10.3758/s13414-013-0595-x

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301. https://doi.org/10.1016/j.actpsy.2011.08.002

Corbett, J. E., Venuti, P., & Melcher, D. (2016). Perceptual Averaging in Individuals with Autism Spectrum Disorder. *Frontiers in Psychology*, *7*, 1735. https://doi.org/10.3389/fpsyg.2016.01735

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012a). An aftereffect of adaptation to mean size. *Visual Cognition*, *20*(2), 211–231. https://doi.org/10.1080/13506285.2012.657261

Cordani, L., Tagliazucchi, E., Vetter, C., Hassemer, C., Roenneberg, T., Stehle, J. H., & Kell, C. A. (2018). Endogenous modulation of human visual cortex activity improves perception at twilight. *Nature Communications*, *9*(1), 1274. https://doi.org/10.1038/s41467-018-03660-8

Cressman, E. K., & Henriques, D. Y. P. (2011). Motor adaptation and proprioceptive recalibration. In *Progress in brain research* (Vol. 191, pp. 91–99). https://doi.org/10.1016/B978-0-444-53752-2.00011-4

de-Wit, L., Machilsen, B., & Putzeys, T. (2010). Predictive Coding and the Neural Response to Predictable Stimuli. *Journal of Neuroscience*, *30*(26).

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Denève, S., Duhamel, J.-R., & Pouget, A. (2007a). Optimal sensorimotor integration in recurrent cortical networks: a neural implementation of Kalman filters. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(21), 5744–5756. https://doi.org/10.1523/JNEUROSCI.3985-06.2007

https://doi.org/10.1523/JNEUROSCI.3985-06.2007

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011). Predictive Context Influences Perceptual Selection during Binocular Rivalry. *Frontiers in Human Neuroscience*, *5*, 166. https://doi.org/10.3389/fnhum.2011.00166

Di Lollo, V., Hogben, J. H., & Dixon, P. (1994). Temporal integration and segregation of brief visual stimuli: Patterns of correlation in time. *Perception & Psychophysics*, *55*(4), 373–386. https://doi.org/10.3758/BF03205295

Einstein, A., & Cowper, A. D. (1926). *INVESTIGATIONS O N THE THEORY .OF ,THE BROWNIAN MOVEMENT R. F ü R T H TRANSLATED BY*. Retrieved from http://users.physik.fu-berlin.de/~kleinert/files/eins_brownian.pdf

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews. Neuroscience*, *9*(4), 292. https://doi.org/10.1038/NRN2258

Feher da Silva, C., Victorino, C. G., Caticha, N., & Baldo, M. V. C. (2017). Exploration and recency as the main proximate causes of probability matching: a reinforcement learning analysis. *Scientific Reports*, *7*(1), 15326. https://doi.org/10.1038/s41598-017-15587-z

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5). https://doi.org/10.1016/j.cub.2014.09.025

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. https://doi.org/10.1038/ncomms2237

Friston, K. (2018). *Does predictive coding have a future?* Retrieved from http://discovery.ucl.ac.uk/10056744/1/Friston_News and views.pdf

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian Inference in Neuroimaging: Applications. https://doi.org/10.1006/nimg.2002.1091

Friston, K., Parr, T., & Zeidman, P. (2018). *Bayesian model reduction*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1805/1805.07092.pdf

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/J.CUB.2017.01.006

Fukushima, K., Fukushima, J., Warabi, T., & Barnes, G. R. (2013). Cognitive processes involved in smooth pursuit eye movements: behavioral evidence, neural substrate and clinical correlation. *Frontiers in Systems Neuroscience*, *7*, 4. https://doi.org/10.3389/fnsys.2013.00004

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*(3), 416–422. https://doi.org/10.1016/J.COGNITION.2008.09.007

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gia Thu, P., & Loc Hung, T. (2003). Bayesian Estimation under Estimation Constraint, *28*(2), 201–207. Retrieved from http://journals.math.ac.vn/acta/pdf/0302201.pdf

Gibson, J. J., & Radner, M. (1937). Adaptation, After Effect and Contrast in the Perception of Tilted Lines.I. Quantatative Studies. Retrieved from https://pdfs.semanticscholar.org/e2be/62e0702894c17780b360d0dbe38a2661c4f3.pdf

Goolsby, B. A., & Suzuki, S. (2001). Understanding priming of color-singleton search: roles of attention at encoding and "retrieval". *Perception & Psychophysics*, *63*(6), 929–944. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11578055

Gordon, N., Koenig-Robert, R., Tsuchiya, N., van Boxtel, J. J., & Hohwy, J. (2017). Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. *ELife*, *6*. https://doi.org/10.7554/eLife.22749

Gregory, R. L. (Richard L. (1970). *Eye and brain : the psychology of seeing*. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=MYgVBgAAQBAJ&oi=fnd&pg=PR5&dq=human+vision+relies+on+light+photons&ots=ACEHinXa9q&sig=7cucxzKLORIFLDc9WuSjxwA2WD0#v=onepage&q=human vision relies on light photons&f=false

Grill-Spector, K., Henson, R., & Martin, A. (2014). Repetition and the brain: neural models of stimulus-specific effects. https://doi.org/10.1016/j.tics.2005.11.006

Grotheer, M. (2016). Can predictive coding explain repetition suppression? *Cortex*, *80*, 113–124. https://doi.org/10.1016/J.CORTEX.2015.11.027

Grover Brown, R., & Hwang, P. Y. C. (2009). Introduction to Random Signals and Applied Kalman Filtering. Retrieved from http://www.biruni.tn/gw_2009_4_3/thumbs/contents-table/163/TM.1632060.pdf

Hall, A. (1945). The British Journal of Opththalmology, September, 1945 Communications-The Origin and purposes of Blinking. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC510520/pdf/brjopthal00688-0001.pdf

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784. https://doi.org/10.1038/29528

Harrison, L. M., Stephan, K. E., Rees, G., & Friston, K. J. (2007). Extra-classical receptive field effects measured in striate cortex with fMRI. *NeuroImage*, *34*(3), 1199–1208. https://doi.org/10.1016/j.neuroimage.2006.10.017

Hatfield, G. (2006). Representation and Constraints;the inverse problem and the structure of visual space. *Acta Psychologica*, *114*, 355–378. https://doi.org/10.1016/j.actpsy.2003.07.003

Haug, A. J. (2012). *Bayesian estimation and tracking : a practical guide*. Wiley.

Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), 1773–1782. https://doi.org/10.1073/pnas.1619788114

Henry, F. M. (1980). Use of Simple Reaction Time in Motor Programming Studies. *Journal of Motor Behavior*, *12*(2), 163–168. https://doi.org/10.1080/00222895.1980.10735216

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*(1), 53–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12927334

Hollingworth, H. L. (1910). The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(17), 461. https://doi.org/10.2307/2012819

Hu, C., Chen, W., Chen, Y., & Liu, D. (2003). Adaptive Kalman Filtering for Vehicle Navigation. *Journal of Global Positioning Systems*, *2*(1), 42–47. Retrieved from https://pdfs.semanticscholar.org/3cc1/16fcda0169c1afdbf26a999d0a337d16908a.pdf

Hua, X., Cheng, Y., Li, Y., Shi, Y., Wang, H., & Qin, Y. (2017). Target detection in sea clutter via weighted averaging filter on the Riemannian manifold. *Aerospace Science and Technology*, *70*, 47–54. https://doi.org/10.1016/j.ast.2017.07.042

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. https://doi.org/10.1002/wcs.142

Hubel David. (1993). Eye, Brain, and Vision. Retrieved February 26, 2018, from http://hubel.med.harvard.edu/book/b8.htm

J. Koehler, D., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, *38*(6), 667–676. https://doi.org/10.3758/MC.38.6.667

Jägerbrand, A. K., & Sjöbergh, J. (2016). Effects of weather conditions, light conditions, and road lighting on vehicle speed. *SpringerPlus*, *5*, 505. https://doi.org/10.1186/s40064-016-2124-6

Jones, J., & Pashler, H. (2009). *Prediction and Retrodiction 1 Is the Mind Inherently Forward-Looking? Comparing Prediction and Retrodiction*. Retrieved from https://escholarship.org/content/qt3hx256zf/qt3hx256zf.pdf

Jones, P. R., & Dekker, T. M. (2018). The development of perceptual averaging: learning what to do, not just how to do it. *Developmental Science*, *21*(3), e12584. https://doi.org/10.1111/desc.12584

Jun, J., & Chong, S. C. (2016). Visual statistical learning of temporal structures at different hierarchical levels. *Attention, Perception, & Psychophysics*, *78*(5), 1308–1323. https://doi.org/10.3758/s13414-016-1104-9

Kafaligonul, H., Breitmeyer, B. G., & Öğmen, H. (2015). Feedforward and feedback processes in vision. *Frontiers in Psychology*, *6*, 279. https://doi.org/10.3389/fpsyg.2015.00279

Kahneman, D. (1968). METHOD, FINDINGS, AND THEORY IN STUDIES OF VISUAL MASKING. *Psychological Bulletin*, *70*(6), 404–425. Retrieved from http://psycnet.apa.org/fulltext/1969-06171-001.pdf

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty : heuristics and biases*. Cambridge University Press.

Kalman, R. E., & Bucy, R. S. (1963). New Results in Linear Filtering and Prediction Theory 1. Retrieved from http://www.eecs.tufts.edu/~khan/Courses/Spring2012/EE194/Lecs/KalmanBucy1961.pdf

Kalman, T., Mi, S., & Bezier, F. (2008). Kalman Filter Applications Predict-Update Equations. *Kalman Filter Tank Filling*, (September 2008), 1–25.

Kanai, R., Knapen, T. H. J., van Ee, R., & Verstraten, F. A. J. (2007). Disruption of implicit perceptual memory by intervening neutral stimuli. *Vision Research*, *47*(20), 2675–2683. https://doi.org/10.1016/J.VISRES.2007.06.016

Kayargadde, V., & Martens, J.-B. (1996a). Estimation of perceived image blur using edge features. *International Journal of Imaging Systems and Technology*, *7*(2), 102–109. https://doi.org/10.1002/(SICI)1098-1098(199622)7:2<102::AID-IMA6>3.0.CO;2-C

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. https://doi.org/10.1007/s10339-007-0170-2

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, *15*(138), 20170792. https://doi.org/10.1098/rsif.2017.0792

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences*, *21*(7), 493–497. https://doi.org/10.1016/J.TICS.2017.04.011

Klapp, S. T. (2010). Comments on the Classic Henry and Rogers (1960) Paper on Its 50th Anniversary. *Research Quarterly for Exercise and Sport*, *81*(1), 108–112. https://doi.org/10.1080/02701367.2010.10599634

Kneissler, J., Drugowitsch, J., Friston, K., & Butz, M. V. (2015). Simultaneous learning and filtering without delusions: a Bayes-optimal combination of Predictive Inference and Adaptive Filtering. *Frontiers in Computational Neuroscience*, *9*, 47. https://doi.org/10.3389/fncom.2015.00047

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7), 5. https://doi.org/10.1167/7.7.5

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. https://doi.org/10.1016/j.tins.2004.10.007

Kö Rding, K. P., & Wolpert, D. M. (2008). Special Issue: Probabilistic models of cognition Bayesian decision theory in sensorimotor control. https://doi.org/10.1016/j.tics.2006.05.003

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Kok, P., De Lange, F. P., Kok, P., & De Lange, F. P. (2017). Predictive Coding in Sensory Cortex. https://doi.org/10.1007/978-1-4939-2236-9_11

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012a). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/J.NEURON.2012.04.034

Kok, P., & Turk-Browne, N. B. (2018). Associative Prediction of Visual Shape in the Hippocampus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *38*(31), 6888–6899. https://doi.org/10.1523/JNEUROSCI.0163-18.2018

Kramer, R. S. S., Jones, A. L., Sharma, D., Glanzer, M., Cunitz, A., Bruin, W. B. de, … Schwarz, N. (2013). Sequential Effects in Judgements of Attractiveness: The Influences of Face Race and Sex. *PLoS ONE*, *8*(12), e82226. https://doi.org/10.1371/journal.pone.0082226

Kristjánsson, Á. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research*, *46*(16), 2554–2570. https://doi.org/10.1016/j.visres.2006.01.015

Kwisthout, J., & Van Rooij, I. (2010). Predictive coding and the Bayesian brain: Intractability hurdles that are yet to be overcome. Retrieved from http://www.socsci.ru.nl/johank/material/cogsci13/paper.pdf

Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 8142–8147. https://doi.org/10.1073/pnas.1500361112

Lee, J., Joshua, M., Medina, J. F., & Lisberger, S. G. (2016). Signal, Noise, and Variation in Neural and Sensory-Motor Latency. *Neuron*, *90*(1), 165–176. https://doi.org/10.1016/J.NEURON.2016.02.012

Levi, D. M., Klein, S. A., & Chen, I. (2005). What is the signal in noise? *Vision Research*, *45*(14), 1835–1846. https://doi.org/10.1016/j.visres.2005.01.020

Li, Y., & Dudman, J. T. (2013). Mice infer probabilistic models for timing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 17154–17159. https://doi.org/10.1073/pnas.1310666110

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*(21). https://doi.org/10.1016/j.cub.2014.09.025

Liberman, A., Zhang, K., & Whitney, D. (2016). Serial dependence promotes object stability during occlusion. *Journal of Vision*, *16*(15), 16. https://doi.org/10.1167/16.15.16

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, *32*(1), 1. https://doi.org/10.2307/1269835

Ly, C., & Doiron, B. (2017a). Noise-enhanced coding in phasic neuron spike trains. *PloS One*, *12*(5), e0176963. https://doi.org/10.1371/journal.pone.0176963

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*(2–3), 81–92. https://doi.org/10.1007/s004260050043

Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE*, *63*(4), 561–580. https://doi.org/10.1109/PROC.1975.9792

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7808275

Maslovat, D., Klapp, S. T., Jagacinski, R. J., & Franks, I. M. (2014). Control of response timing occurs during the simple reaction time interval but on-line for choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2005–2021. https://doi.org/10.1037/a0037522

Melcher, D. (2011). Visual stability. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1564), 468–475. https://doi.org/10.1098/RSTB.2010.0277

Moors, P., Stein, T., Wagemans, J., & Ee, R. van. (2015). Serial correlations in Continuous Flash Suppression. *Neuroscience of Consciousness*, *2015*(1), niv010. https://doi.org/10.1093/nc/niv010

Newell, B. R., Koehler, D. J., James, G., Rakow, T., & van Ravenzwaaij, D. (2013). Probability matching in risky choice: The interplay of feedback and strategy availability. *Memory & Cognition*, *41*(3), 329–338. https://doi.org/10.3758/s13421-012-0268-3

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Niv, Y., Joel, D., Meilijson, I., & Ruppin, E. (2002). Evolution of Reinforcement Learning in Uncertain Environments: A Simple Explanation for Complex Foraging Behaviors. *Adaptive Behavior*, *10*(1), 5–24. https://doi.org/10.1177/10597123020101001

Nuding, U., Kalla, R., Muggleton, N. G., Bu¨ttnerbu¨ttner, U., Walsh, V., & Glasauer, S. (2009). TMS Evidence for Smooth Pursuit Gain Control by the Frontal Eye Fields. *Cerebral Cortex May*, *19*, 1144–1150. https://doi.org/10.1093/cercor/bhn162

Nummiaro, K., Koller-Meier, E., & Van Gool, L. (2003). An adaptive color-based particle filter. *Image and Vision Computing*, *21*(1), 99–110. https://doi.org/10.1016/S0262-8856(02)00129-4

O'Shaughnessy, D. (1988). Linear predictive coding. *IEEE Potentials*, *7*(1), 29–32. https://doi.org/10.1109/45.1890

Olivers, C. N. L., & Meeter, M. (2008). Feature priming in visual search does not depend on the dimensional context. *Visual Cognition*, *16*(6), 785–803. https://doi.org/10.1080/13506280701428658

Ono, S., & Mustari, M. J. (2008). Role of the MST-DLPN pathway in smooth pursuit adaptation. *Progress in Brain Research*, *171*, 161–165. https://doi.org/10.1016/S0079-6123(08)00621-3

Orfanidis, S. J. (1988). *Optimum signal processing : an introduction*. McGraww-Hill.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). *Compulsory averaging of crowded orientation signals in human vision*. Retrieved from http://neurosci.nature.com

Parr, T., Rees, G., & Friston, K. J. (2018). Computational Neuropsychology and Bayesian Inference. *Frontiers in Human Neuroscience*, *12*, 61. https://doi.org/10.3389/fnhum.2018.00061

PEARSON, K. (1905). The Problem of the Random Walk. *Nature*, *72*(1865), 294–294. https://doi.org/10.1038/072294b0

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Praveena, K., & Ravikumar, A. (2013). Design of Inertial Navigation System using Kalman Filter. *International Journal of Engineering Inventions*, *2*(12), 2278–7461. Retrieved from www.ijeijournal.com

Rao, R. P. N. (1999). *An optimal estimation approach to visual perception and learning*. *Vision Research* (Vol. 39). Retrieved from https://homes.cs.washington.edu/~rao/vr99.pdf

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *2*(1), 79–87. https://doi.org/10.1038/4580

Robbins, D., & Warner, P. L. (1973). Individual organism probability matching with rats in a two-choice task. *Bulletin of the Psychonomic Society*, *2*(6), 405–407. https://doi.org/10.3758/BF03334429

Robinson, D. A. (1965). The Mechanics of Human Smooth Pursuit Eye Movement .*J. Physiol* (Vol. 180). Retrieved from https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1965.sp007718

Rønn-Nielsen, A., & Hansen, E. (2014). *Conditioning and Markov properties*. Retrieved from http://web.math.ku.dk/noter/filer/beting.pdf

Rudin, C., Letham, B., Salleb-Aouissi, A., Kogan Sourcetone, E., & Madigan, D. (2008). *Sequential Event Prediction with Association Rules*. Retrieved from https://pdfs.semanticscholar.org/537e/5e9cf12c118c85ee665f22ea51d2cff53355.pdf

Rukhin, A. L. (2009). Weighted means statistics in interlaboratory studies. *Metrologia*, *46*, 323–331. https://doi.org/10.1088/0026-1394/46/3/021

Saijo, N., & Gomi, H. (2012). Effect of visuomotor-map uncertainty on visuomotor adaptation. *Journal of Neurophysiology*, *107*(6), 1576–1585. https://doi.org/10.1152/jn.00204.2011

Sanz, I., Calvo, M. L., Chevalier, M., & Lakshminarayanan, V. (2001). Perception of High-Contrast Blurred Edges. *Journal of Visual Communication and Image Representation*, *12*(3), 240–254. https://doi.org/10.1006/JVCI.2001.0473

Scheidt, R. A., Dingwell, J. B., & Mussa-Ivaldi, F. A. (2001). Learning to Move Amid Uncertainty. *Journal of Neurophysiology*, *86*(2), 971–985. https://doi.org/10.1152/jn.2001.86.2.971

Schmitt, C., Klingenhoefer, S., & Bremmer, F. (2018). Preattentive and Predictive Processing of Visual Motion. *Scientific Reports*, *8*(1), 12399. https://doi.org/10.1038/s41598-018-30832-9

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Searle, J. (2015). *Seeing Things as They Are*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199385157.001.0001

Seth, A. K. (2014). The Cybernetic Bayesian Brain. https://doi.org/10.15502/9783958570108

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2012). Unsupervised Learning of Generative and Discriminative Weights Encoding Elementary Image Components in a Predictive Coding Model of Cortical Function. *Neural Computation*, *24*(1), 60–103. https://doi.org/10.1162/NECO_a_00222

Spratling, M. W. (2015). A review of predictive coding algorithms. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2015.11.003

Srinivasan, M. V, Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459. https://doi.org/10.1098/RSPB.1982.0085

St. John-Saaltink, E., Kok, P., Lau, H. C., de Lange, F. P., St John-Saaltink, E., Kok, P., … de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *36*(23), 6186–6192. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Strong, S. P., Koberle, R., De Ruyter Van Steveninck, R. R., & Bialek, W. (1998). *Entropy and Information in Neural Spike Trains*. Retrieved from http://www.stat.columbia.edu/~liam/teaching/neurostat-spr12/papers/bialek-et-al/strong+al_98a.pdf

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., & Koechlin, E. (2008). A Neural Representation of Prior Information during Perceptual Inference. *Neuron*, *59*(2), 336–347. https://doi.org/10.1016/j.neuron.2008.05.021

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Frontiers in Human Neuroscience*, *5*, 67. https://doi.org/10.3389/fnhum.2011.00067

Swain, P. S., & Longtin, A. (2006). Noise in genetic and neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *16*(2), 026101. https://doi.org/10.1063/1.2213613

Swift, D. J. (2013). Temporal integration in vision: Masking can aid detection. *Attention, Perception, & Psychophysics*, *75*(3), 481–490. https://doi.org/10.3758/s13414-012-0418-5

Takahashi, S., & Miwa, S. (1994). Suppression of radar clutter with a weighted moving average filter. *Electronics and Communications in Japan (Part I: Communications)*, *77*(1), 57–65. https://doi.org/10.1002/ecja.4410770106

Taubert, J., & Alais, D. (2016). Serial dependence in face attractiveness judgements tolerates rotations around the yaw axis but not the roll axis. *Visual Cognition*, *24*(2), 103–114. https://doi.org/10.1080/13506285.2016.1196803

Taubert, J., Alais, D., Burr, D., Gepshtein, S., Lesmes, L. A., Albright, T. D., … Perrett, D. I. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*, 32239. https://doi.org/10.1038/srep32239

Tavassoli, A., & Ringach, D. L. (2009). Dynamics of smooth pursuit maintenance. *Journal of Neurophysiology*, *102*(1), 110–118. https://doi.org/10.1152/jn.91320.2008

Taylor, J. R. (John R. (1997). *An introduction to error analysis : the study of uncertainties in physical measurements*. University Science Books. Retrieved from https://books.google.co.uk/books?id=giFQcZub80oC&pg=PA94&redir_esc=y

Thier, P., & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. https://doi.org/10.1016/j.conb.2005.10.013

Thornton, C. (2014). *Experiments in Predictive Coding* *. Retrieved from http://www.lorentzcenter.nl/lc/web/2014/627/presentations/Chris Thornton : Experiments in Predictive Coding.pdf

Todorov, E. (2004). Optimality principles in sensorimotor control (review). *Nature Neuroscience*, *7*(9), 907. https://doi.org/10.1038/NN1309

Turk-Browne, N. B. (2012). Statistical Learning and Its Consequences (pp. 117–146). https://doi.org/10.1007/978-1-4614-4794-8_6

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Tuzlukov, V. P. (V͡iacheslav P. (2002). *Signal processing noise*. CRC Press.

Unturbe, J., & Corominas, J. (2007). Probability matching involves rule-generating ability: A neuropsychological mechanism dealing with probabilities. *Neuropsychology*, *21*(5), 621–630. https://doi.org/10.1037/0894-4105.21.5.621

Urteaga, I., & Wiggins, C. H. (2018). *Bayesian bandits: balancing the exploration-exploitation tradeoff via double sampling*. Retrieved from https://arxiv.org/pdf/1709.03162.pdf

Van Der Helm, P. A. (2016). Structural coding versus free-energy predictive coding. https://doi.org/10.3758/s13423-015-0938-9

Walsh Flaherty N D Stanley Coren, Eugen. A. (1974). Reactiom Time as a Measure of the Effect of Selective Attention and Motor Skills (Vol. 39). Retrieved from https://journals.sagepub.com/doi/pdf/10.2466/pms.1974.39.2.755

Warrant, E. J., & Johnsen, S. (2013). Vision and the light environment. *Current Biology*, *23*(22), R990–R994. https://doi.org/10.1016/J.CUB.2013.10.019

Warren, W. H. (2012). Does this computational theory solve the right problem? Marr, Gibson, and the goal of vision. *Perception*, *41*(9), 1053–1060. https://doi.org/10.1068/p7327

Waugh, S. J., & Levi, D. M. (1993). Visibility, timing and vernier acuity. *Vision Research*, *33*(4), 505–526. https://doi.org/10.1016/0042-6989(93)90255-U

Wei, K., & Koerding, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Frontiers in Computational Neuroscience*, *4*, 11. https://doi.org/10.3389/fncom.2010.00011

Wei, K., & Körding, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Frontiers in Computational Neuroscience*, *4*, 11. https://doi.org/10.3389/fncom.2010.00011

Wei, K., Wert, D., & Körding, K. (2010). The Nervous System Uses Nonspecific Motor Learning in Response to Random Perturbations of Varying Nature. *Journal of Neurophysiology*, *104*(6), 3053–3063. https://doi.org/10.1152/jn.01025.2009

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. *In Practice*, *7*(1), 1–16. https://doi.org/10.1.1.117.6808

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9635206

Wolfensteller, U., & Ruge, H. (2011). On the timescale of stimulus-based action–effect learning. *Quarterly Journal of Experimental Psychology*, *64*(7), 1273–1289. https://doi.org/10.1080/17470218.2010.546417

Wolpert, D., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. *Human Movement Science*, *26*(4), 511–524. https://doi.org/10.1016/j.humov.2007.05.005

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science, New Series*, *269*(29), 1880–1882. Retrieved from http://links.jstor.org/sici?sici=0036-8075%2819950929%293%3A269%3A5232%3C1880%3AAIMFSI%3E2.0.CO%3B2-Q

Yoshimoto, S., Takeuchi, T., A., B. D., J., B. O., H., B. D., L., B., … K., Y. (2013). Visual motion priming reveals why motion perception deteriorates during mesopic vision. *Journal of Vision*, *13*(8), 8–8. https://doi.org/10.1167/13.8.8

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zambrano, D., Falotico, E., Manfredi, L., & Laschi, C. (2010). A model of the smooth pursuit eye movement with prediction and learning. *Applied Bionics and Biomechanics*, *7*(2), 109–118. https://doi.org/10.1080/11762321003760944

Zarchan, P., & Musoff, H. (2000). *Fundamentals of Kalman filtering : a practical approach*. American Institute of Aeronautics and Astronautics. Retrieved from https://books.google.co.uk/books?id=AQxRAAAAMAAJ&redir_esc=y

Zeil, J., Boeddeker, N., & Hemmi, J. M. (2008). Vision and the organization of behaviour. *Current Biology*, *18*(8), R320–R323. https://doi.org/10.1016/J.CUB.2008.02.017

Zetterberg, L.-H. (1958). Detection of moving radar targets in clutter. *Information and Control*, *1*(4),

314–333. https://doi.org/10.1016/S0019-9958(58)90209-2

Zumbahlen, H., & Zumbahlen, H. (2008). Analog Filters. *Linear Circuit Design Handbook*, 581–679. https://doi.org/10.1016/B978-0-7506-8703-4.00008-0

Zylberberg, J., Pouget, A., Latham, P. E., & Shea-Brown, E. (2017). Robust information propagation through noisy neural circuits. *PLOS Computational Biology*, *13*(4), e1005497. https://doi.org/10.1371/journal.pcbi.1005497

8. Appendices

8.1 Ethical approval

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

17 November 2015

Dear Justin and Fraser

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 3rd November 2015; the following documents have been reviewed:

1. Ethical Application Form
2. Experimental Procedure Outline
3. Advertisement
4. Participant Information Sheet
5. Consent Form
6. Debriefing Form
7. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS11816	**Approved on:**	16/11/2015	**Approval Expiry:**	16/11/2018
Project Title:	The Kalman Brain? How does the brain use prediction to regulate temporal aspects of the world.				
Researcher:	Dr Justin Ales and Fraser Aitken				
Supervisor:	Dr Justin Ales				

Approval is awarded for three years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the 3 three year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.2 Work related to this thesis

Title: Adaptive serial dependencies in visual perception. Venue/date: European conference of visual perception. Frei Universitat, Berlin, German, Aug 2017

Title: Quantifying sensory noise in serial dependence experiments, Venue: Vision sciences society conference. Trade winds resort, Florida, USA , May 2018

The following paper is currently under way.

Title-Adaptive serial dependence in visual perception.

8.3 Analysis and stimulus code can be found at

https://github.com/aleslab/analysis